

II. FACTUAL BACKGROUND¹

The DOC operates eighteen correctional facilities throughout Connecticut. See Facilities Page, Mot. for Class Cert., Doc. No. 57, Ex. 5. The DOC employs COs who are responsible for the confinement, safety, control, and monitoring of inmates in these correctional facilities, and for the security of the facilities. Defendant's Local Rule 56(a)(1) Statement ("Def.'s 56(a)(1)"), ¶ 2. In order to obtain a position as a CO within one of these DOC facilities, a candidate must possess a high school diploma or G.E.D., be at least twenty-one years of age, be in good health, have a good educational or work record, and be free from felony convictions. Plaintiff's Corrected Local Rule 56(a)(1) Statement ("Pl.'s 56(a)(1)"), Ex. 1 (2004 correction officer job posting), Ex. 2 (2006 correction officer job posting). Applicants who meet these criteria first take a written examination. Pl.'s 56(a)(1), ¶ 2. If an applicant passes the written exam, the applicant then takes a physical fitness test ("PFT"). Pl.'s 56(a)(1), ¶ 4. Only if an applicant passes both the written and physical examinations is his or her application turned over to the DOC for interviews, background checks, and medical examinations. Pl.'s 56(a)(1), ¶ 8, Exs. 1, 2.

The physical fitness test consists of four parts. Def.'s 56(a)(1), ¶ 23. Failing any single part causes the candidate to fail the entire test. Id. In 2004 and 2006, the physical fitness test included: (1) a sit and reach test; (2) a one-minute sit-up test; (3) a one-minute push-up test; and (4) a timed 1.5 mile run. Id. The minimum standards for candidates varied by gender and age. For example, a female candidate in the 21-29 age group was required to complete 1.5 mile run in 14 minutes, 49 seconds. Pl.'s

¹ Unless otherwise noted, no genuine dispute exists over the following facts.

56(a)(1), Ex. 4 (physical fitness score sheet). The corresponding standard for a male candidate in the 20-29 age group was 12 minutes, 25 seconds. Id. These standards are set to the 40th percentile of performance for each age/gender cohort, as calculated by the Cooper Institute. Pl.'s 56(a)(1), ¶¶ 17, 19. The passing time for each age and gender combination was:

Gender / Age	21-29	30-39	40-49	50+
Male	12:25	12:51	13:46	14:54
Female	14:49	15:25	16:12	17:14

Pl.'s 56(a)(1), Physical Fitness Score Sheet, Ex. 4.

The 1.5 mile timed run was administered as part of the PFT in October 2004 and June 2006; a make-up exam with a 1.5 mile timed run was also administered in October 2006. Pl.'s 56(a)(1), ¶ 13; Defendant's Local Rule 56(a)(2) Statement (in Response to Plaintiff's Corrected Local Rule 56(a)(1) Statement) ("Def.'s 56(a)(2)"), ¶ 13; Def.'s 56(a)(1), ¶¶ 34-35. For the October 2006 administration of the 1.5 mile run, applicants were either individuals who failed the June 2006 administration or individuals who opted to train for several months before attempting the timed run. Def.'s 56(a)(1), ¶¶ 34-35. For each of these tests, the DOC used the Cooper Institute 40th percentile marks as minimum standards. Pl.'s 56(a)(1), ¶¶ 17-19.

In mid-2004, Cherie Easterling, a resident of Bloomfield, Connecticut, applied for employment as a Correction Officer with the DOC. Pl.'s 56(a)(1), ¶ 1. Easterling took and passed the written exam on August 25, 2004. Pl.'s 56(a)(1), ¶ 3. In October 2004, Easterling took the physical fitness test. Pl.'s 56(a)(1), ¶¶ 5-6. Although Easterling passed the 'sit and reach test,' the 'one-minute sit-up test,' and the 'one-minute push-up test,' she failed the 1.5 mile run portion of the test. Id. As a result of failing the 1.5 mile

run test, Easterling was precluded from advancing in the DOC selection process. Pl.'s 56(a)(1), ¶ 7.

A. Statistical Evidence Related to Disparate Impact

1. Gender Disparity in Percentage of CO Applicants Passing the 1.5 Mile Test

On each of these three occasions that the 1.5 mile test was administered, women passed the test at a substantially lower percentage than men.

In 2004, 148 women were eligible to participate in the 1.5 mile run portion of the PFT. Pl.'s 56(a)(1), ¶ 36, Ex. 7. Ninety-three (93) of those women, or 62.8%, passed the 1.5 mile run component of the PFT. Id. Seven hundred forty (740) men completed the 1.5 mile run, and 607 passed, for a passage rate of 82.0%. Id. The ratio of the female passage rate to the male passage rate was 76.6%.

In June 2006, 172 women took the 1.5 mile run portion of the PFT. Pl.'s 56(a)(1), ¶ 38, Ex. 8. One hundred one (101) of those women, or 58.7%, passed the 1.5 mile run component of the PFT. Id. Eight hundred twenty-six (826) men completed the 1.5 mile run, and 690 passed, for a passage rate of 83.5%. Id. The ratio of the female passage rate to the male passage rate was 70.3%.

In October 2006, the DOC offered a make-up administration of the PFT for applicants who had failed the June 2006 PFT and applicants who opted to train for additional time before attempting the PFT. Pl.'s 56(a)(1), ¶ 40, Ex. 9. That month, seventy-eight (78) women took the 1.5 mile run portion of the PFT. Twenty-seven (27) of those women, or 34.6%, passed the 1.5 mile run component of the PFT. Id. Two hundred fifty-eight (258) men completed the 1.5 mile run, and 137 passed, for a

passage rate of 53.1%. Id. The ratio of the female passage rate to the male passage rate was 65.2%.

Over the three administrations of the PFT at issue in this lawsuit, 398 women participated in the 1.5 mile run portion of the PFT, and 221 women passed, for a passage rate of 55.5%. For the same three administrations, 1,824 men participated in the 1.5 mile run portion of the PFT, and 1,434 men passed, for a passage rate of 78.6%. Overall, the ratio of the female passage rate to the male passage rate was 70.6%.

The plaintiff's statistical expert, Dr. Alexander Vekker, opined that the 2004, June 2006, and September 2006 administrations of the 1.5 mile run component of the PFT each yielded statistically significant gender disparities in outcomes. Pl.'s 56(a)(1), Ex. 24, at 1. Using the statistical method known as "Fisher's Exact Test,"² Dr. Vekker estimated that the probability of the gender difference observed for the 2004 administration of the 1.5 mile run occurring by chance was less than one in 1.5 million, or 4.9 standard deviations. Id. Dr. Vekker estimated that the probability of the gender difference observed for the June 2006 administration occurring by chance was less than one in 145 billion, or 6 standard deviations. Id. Finally, Dr. Vekker estimated that the probability of the gender difference observed for the September 2006 administration occurring by chance was less than one in 64,000, or 4.2 standard deviations. Id. Dr.

² Fisher's Exact Test is "a statistical method commonly used in employment discrimination cases." Schanzer v. United Technologies Corp., 120 F. Supp. 2d 200, 204 (D. Conn. 2000) (citing Victory v. Hewlett-Packard Co., 34 F. Supp. 2d 809 (E.D.N.Y. 1999)). This method can be used to test the hypothesis that whether an applicant passes the 1.5 mile run is statistically independent of his or her sex. See Michael O. Finkelstein & Bruce A. Levin, STATISTICS FOR LAWYERS 154-57 (2001).

Vekker estimated that 71 additional women would have passed the 1.5 mile run over the three administrations of the PFT, if the run had been gender neutral. Id.

The DOC admits that female CO applicants failed the 1.5 mile run component of the PFT at a higher rate than male CO applicants. Def.'s 56(a)(2), ¶ 42. In its Local Rule 56(a)(2) Statement, the DOC argues that the disparity "did not always rise to [the] level of adverse impact." Id. However, the DOC does not offer any evidence in support of this argument. The defendant's sole expert on statistical disparities, Dr. Harold Brull, concluded that the 1.5 mile run event had an adverse impact on female CO applicants. Pl.'s 56(a)(1), Deposition of Harold Brull, Ex. 19, at 63.³ In his expert report, Dr. Brull concluded that "it is obvious, upon even the most cursory review of the applicant test data, that female applicants for the position of Correction Officer have a slightly more difficult time completing the event than their male counterparts." Pl.'s 56(a)(1), Report of Harold P. Brull, Ex. 26, at 3. Dr. Brull also "checked the math" in Dr. Vekker's report and concluded that Dr. Vekker's calculations of the probability of the observed disparity occurring by chance were accurate. Pl.'s 56(a)(1), Deposition of Harold Brull, Ex. 19, at 63-64.

2. Gender Disparity When Positions Other than CO Are Considered

DOC has presented evidence that when the results of the 1.5 mile run are pooled for the positions of State Trooper Trainee, Public Safety Trainee, and CO, the observed disparity between men and women declines. Specifically, in 2004, women applying for

³ The DOC designated Dr. Harold Brull as its expert on statistical disparities in the 1.5 mile timed run. See Def.'s Designation of Expert Concerning Statistical Disparities, Doc. No. 82. DOC also designated Dr. Brull as an expert on Business Necessity and Job Relatedness. See Def.'s Designation of Expert Concerning Business Necessity and Job Relatedness, Doc. No. 90.

these three positions passed the 1.5 mile run at a rate of 75.5%. Men passed at a rate of 86.5%. The ratio of women to men passing the test was 87.3%. Def.'s 56(a)(1), ¶ 24. In 2006, 72.6% of women applying for these positions passed the 1.5 mile run, while 91.0% of men passed. The ratio of women to men passing the test was 79.8%. Def.'s 56(a)(1), ¶ 26.⁴

3. Why Female Applicants Disproportionately Failed 1.5 Mile Run

A genuine dispute exists over why female CO applicants failed the timed run at a higher rate than male CO applicants. The use of normed cut scores should yield comparable pass rates for each age and gender cohort. The plaintiff has presented expert testimony that the Cooper percentile rankings used by the DOC were based on measurements from a sample of women who possessed a higher level of fitness than the overall female population. See Pl.'s 56(a)(1), ¶ 31. Ex. 15, at 13-14; Ex. 22, at 7-8.⁵

The DOC argues instead that its efforts to recruit racial minorities, who tend to have lower levels of cardiovascular fitness, produced a female applicant pool that was less likely to pass the 1.5 mile run than the overall pool of qualified workers. Def.'s Mem. in Opp., at 13-15. The DOC has not offered any evidence that its recruitment efforts in the Hispanic and African-American communities were specifically targeted at women in those communities. See, e.g., Def.'s Supplemental 56(a)(1), ¶ 52 (stating that the DOC "did quite a bit of outreach to the Hispanic community"). In addition, the DOC has not presented statistical evidence that minority women are significantly less

⁴ The court infers that the DOC combined the data for the June 2006 and October 2006 administrations of the run exam.

⁵ The defendant denies this conclusion on the grounds that the expert's opinion is not based on admissible evidence. Def.'s 56(a)(2), ¶ 31.

likely to pass the 1.5 mile run than minority men. Without such a showing, no reasonably jury could conclude that an effort to recruit minority applicants of both genders would alter the applicant pool in a manner that creates a disparity in outcomes by gender. The DOC has thus failed to present evidence that its own recruitment efforts caused men and women to pass the timed run at different rates.

B. Evidence Regarding “Business Necessity” and “Job-Relatedness”

The parties agree that the 1.5 mile run is a test that measures an individual’s aerobic capacity. Def.’s 56(a)(1), ¶ 12; Pl.’s 56(a)(2), ¶ 12. The defendant has presented no evidence showing the timed 1.5 mile run to be predictive of who can perform the essential physical functions of the job of a CO. All three of the defendant’s experts admit that they have not empirically demonstrated that the cut scores used by DOC for the 1.5 mile run reliably predict an individual’s performance on particular job tasks as a CO.

Dr. Pamela Libby is the Director of Statewide Human Resources Management in the Department of Administrative Services and one of the defendant’s experts on the issue of the job relatedness of the 1.5 mile run. Def.’s Designation of Expert Concerning Business Necessity and Job Relatedness, Doc. No. 91; Pl.’s 56(a)(1), Ex. 27, at 5. Dr. Libby testified at her deposition on July 15, 2010, that she could not correlate the 40th percentile cut scores that were used on the 1.5 mile run test with the minimum level of aerobic capacity necessary to perform the CO position. Pl.’s 56(a)(1), Ex. 25, at 116.

The DOC also designated Dr. Harold Brull as an expert on Business Necessity and Job Relatedness. See Def.’s Designation of Expert Concerning Business

Necessity and Job Relatedness, Doc. No. 90. At his deposition, Dr. Brull stated that he had not attempted to determine the minimum level of aerobic capacity necessary to successfully perform the CO position “with any level of exactness through a statistical test or empirical validation.” Pl.’s 56(a)(1), ¶ 21, Ex. 19, at 87. Dr. Brull also testified that he had not observed any statistically significant correlation between aerobic capacity and CO performance. Pl.’s 56(a)(1), ¶ 23, Ex. 19, at 125. Instead, Dr. Brull opined that he believed cut scores set at the 40th percentile “seems a reasonable standard and not an excessive one.” Pl.’s 56(a)(1), ¶ 21, Ex. 19, at 87. Dr. Brull concluded that “it’s reasonable to assume [that] higher levels of fitness correlate with more positive outcomes and fewer negative outcomes,” but did not provide any evidence in support of that assumption. See Pl.’s 56(a)(1), ¶ 24, Ex. 19, at 116.

The DOC’s third expert on the issue of job-relatedness concurred. The DOC designated Dr. Martin Anderson, Commissioner of Connecticut’s Department of Administrative Services, as an expert on business necessity and job relatedness. See Def.’s Designation of Expert Concerning Business Necessity and Job Relatedness, Doc. No. 91. At his deposition, Dr. Anderson testified that DAS had not “measured the specific cardiovascular capacity required” for the tasks performed by a CO that involve “endurance” and “cardiovascular capacity.” Pl.’s 56(a)(1), ¶ 22, Ex. 14, at 13-14.

The defendant states that it selected cut times at the 40th percentile because the Cooper Institute labels performance at that level as “fair” rather than “poor.” See Def.’s 56(a)(1), ¶¶ 10, 13. However, the Cooper Institute warns its consumers not to use the thresholds at a given percentile as a hiring standard. Pl.’s 56(a)(1), ¶ 29, Ex. 20, at 3. The Cooper Institute states that “[u]sing percentile rankings of the Cooper norms for

standards is not defensible” because the “percentile rankings do not predict the ability to do the job and do not demonstrate criterion validity.”⁶ Id. The Cooper Institute also states that “percentile scores (whether age and gender norms or single norms) have no validity data for predicting who can and who cannot do the job.” Pl.’s 56(a)(1), ¶ 30, Ex. 21, at 9.

C. DOC’s Subsequent Replacement of the Timed 1.5 Mile Run

In 2007, the DOC modified the physical fitness component of its selection process. Instead of a timed 1.5 mile run, the DOC is now piloting a 300-meter run test. Def.’s 56(a)(1), ¶ 38. In order to graduate from the training academy, cadets must pass a 300 meter run test at the 40th percentile for their age and gender group. Pl.’s 56(a)(1), ¶ 54.

III. STANDARD OF REVIEW

A motion for summary judgment “may properly be granted . . . only where there is no genuine issue of material fact to be tried, and the facts as to which there is no such issue warrant judgment for the moving party as a matter of law.” In re Dana Corp., 574 F.3d 129, 151 (2d Cir. 2009). Thus, the role of a district court in considering such a motion “is not to resolve disputed questions of fact but only to determine whether, as to any material issue, a genuine factual dispute exists.” Id. In making this determination, the trial court must resolve all ambiguities and draw all inferences in favor of the party against whom summary judgment is sought. See Fed R. Civ. P. 56(c); Loeffler v. Staten Island Univ. Hosp., 582 F.3d 268, 274 (2d Cir. 2009).

⁶ Criterion validity is demonstrated by identifying criteria that indicate successful job performance and then correlating test scores with those identified criteria. See Guardian Assoc. of New York City Police Dep’t, Inc. v. Civil Service Commission of the City of New York, 633 F.2d 232, 242, n.16 (2d Cir. 1980).

“[T]he moving party bears the burden of showing that he or she is entitled to summary judgment.” United Transp. Union v. National R.R. Passenger Corp., 588 F.3d 805, 809 (2d Cir. 2009). Once the moving party has satisfied that burden, in order to defeat the motion, “the party opposing summary judgment . . . must set forth ‘specific facts’ demonstrating that there is ‘a genuine issue for trial.’ ” Wright v. Goord, 554 F.3d 255, 266 (2d Cir. 2009) (quoting Fed. R. Civ. P. 56(e)). “A dispute about a ‘genuine issue’ exists for summary judgment purposes where the evidence is such that a reasonable jury could decide in the non-movant's favor.” Beyer v. County of Nassau, 524 F.3d 160, 163 (2d Cir. 2008) (quoting Guilbert v. Gardner, 480 F.3d 140, 145 (2d Cir. 2007)); see also Havey v. Homebound Mortgage, Inc., 547 F.3d 158, 163 (2d Cir. 2008) (stating that a non-moving party must point to more than a mere “ ‘scintilla’ ” of evidence in order to defeat a motion for summary judgment) (citing Anderson v. Liberty Lobby, Inc., 477 U.S. 242, 252 (1986)). In this case, there are no genuine issues of material fact to be tried, and the court is able to adjudicate the Motions for Summary Judgment as a matter of law.

IV. DISCUSSION

Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e et seq., as amended, prohibits intentional discrimination on the basis of race, color, religion, sex, or national origin, as well as employment practices “that are not intended to discriminate but in fact have a disproportionately adverse impact on minorities (known as ‘disparate impact’).” Ricci v. DeStefano, 129 S. Ct. 2658, 2672 (2009). The Civil Rights Act of 1964 did not expressly prohibit policies yielding a “disparate impact,” but the Supreme Court interpreted the Act in Griggs v. Duke Power Co., 401 U.S. 424 (1971), to “proscribe[] not only overt discrimination but also practices that are fair in form, but discriminatory in operation.” Id. at 431. The Griggs Court explained that “good intent or 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.” Id. at 432. “The touchstone is business necessity. If an employment practice which operates to exclude [minorities] cannot be shown to be related to job performance, the practice is prohibited.” Id. at 431. In 1991, Congress codified the prohibition on disparate impact discrimination that was established in Griggs. See Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (1992) (codified in scattered sections of 2 U.S.C., 29 U.S.C., and 42 U.S.C.).

Under the disparate impact statute, courts conduct “a three-part analysis involving shifting evidentiary burdens.” Gulino v. New York State Education Dep’t, 460 F.3d 361, 382 (2d Cir. 2006). First, “a plaintiff establishes a prima facie violation by showing that an employer uses ‘a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin.’” Ricci, 129

S. Ct. at 2673 (citing 42 U.S.C. § 2000e-2(k)(1)(A)(i)). Second, “the defendant has two avenues of rebuttal” after “the plaintiff establishes a prima facie case of disparate impact discrimination.” Gulino, 460 F.3d at 382. The employer “may directly attack plaintiff’s statistical proof by pointing out deficiencies in the data or fallacies in the analysis.” Id. The employer may also overturn the plaintiff’s prima facie showing by proving that “the employment practice that causes a disparate impact . . . is job related for the position in question and consistent with business necessity.” 42 U.S.C. § 2000e-2(k)(1)(A)(i). Third, if the employer is successful in rebutting the plaintiff’s prima facie case, the “plaintiff may still succeed by showing that the employer refuses to adopt an available alternative employment practice that has less disparate impact and serves the employer’s legitimate needs.” Ricci, 129 S. Ct. at 2673 (citing 42 U.S.C. § 2000e-2(k)(1)(A)(ii)).

A. Plaintiff’s Prima Facie Case of Disparate Impact

1. The Employment Practice Causing the Disparity

“To make out a prima facie case of disparate impact, a plaintiff must ‘(1) identify a policy or practice, (2) demonstrate that a disparity exists, and (3) establish a causal relationship between the two.’” Malave v. Potter, 320 F.3d 321, 325 (2d Cir. 2003) (quoting Robinson v. Metro-North Commuter R.R. Co., 267 F.3d 147, 160 (2d Cir. 2001)). “[T]he plaintiff must first identify the employment practice allegedly responsible for the disparities.” Gulino, 460 F.3d at 382 (citing 42 U.S.C. § 2000e-2(k)(1)(B)(i)). In this case, the plaintiff has identified the timed 1.5 mile run segment of the PFT as the employment practice causing the disparate impact. Specifically, the plaintiff alleges that

female CO applicants disproportionately fail the 1.5 mile run test, leaving an inadequate percentage of women in the remaining candidate pool.

2. Plaintiff's Statistical Evidence of Disparity

A plaintiff may not "rely on bottom line numbers in an employer's workforce" to establish a prima facie case of disparate impact. Smith v. Xerox Corp., 196 F.3d 358, 365 (2d Cir. 1999) (citation omitted) (overruled on other grounds by Meacham v. Knolls Atomic Laboratory, 461 F.3d 134 (2d Cir. 2006)). Instead, the plaintiff must produce statistical evidence showing that the identified employment practice "causes a disparate impact." Gulino, 460 F.3d at 382. Here, the plaintiff has presented evidence that women passed the 1.5 mile timed run at a lower rate than men, and this has resulted in a pool of candidates for the CO position that is disproportionately male. Over the three administrations of the PFT at issue in this lawsuit, 55.5% of women passed the 1.5 mile run, whereas 78.6% of men passed the 1.5 mile run. The ratio of the female passage rate to the male passage rate was 70.6%.

To establish a causal relationship between the employment practice and this observed disparity, the plaintiff's statistical evidence must reveal a disparity that is "sufficiently substantial to raise an inference of causation." Xerox Corp., 196 F.3d at 365 (citations omitted); Waisome v. Port Authority of New York and New Jersey, 948 F.2d 1370, 1375 (2d Cir. 1991). Courts in the Second Circuit have embraced two principal methods for determining whether a plaintiff has presented statistical data sufficient to raise an inference of causation. See Xerox Corp., 196 F.3d at 365. The first method is the "four-fifths rule" found in the U.S. Equal Employment Opportunity Commission (EEOC) Uniform Guidelines on Employment Selection Procedures:

A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded . . . as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from that group.

29 C.F.R. § 1607.4(D) (2011). The four-fifths rule “is not binding on courts, and is merely a ‘rule of thumb’ to be considered in appropriate circumstances.” EEOC v. Joint Apprenticeship Committee of the Joint Industry Bd. of the Electrical Industry, 186 F.3d 110, 118 (2d Cir. 1999). Keeping in mind that a selection rate below the four-fifths standard does not automatically establish a causal relationship, the plaintiff has demonstrated that female performance on the 1.5 mile run fell below the four-fifths threshold.

The second method for determining whether a plaintiff has presented statistical data sufficient to raise an inference of causation is a statistical test that measures the probability that observed differences in outcomes between two groups are attributable to random variation. Xerox Corp., 196 F.3d at 366. The Second Circuit generally classifies a disparity as statistically significant if the observed disparity exceeds two standard deviations. See Malave v. Potter, 320 F.3d 321, 327 (2d Cir. 2003). “If an obtained result varies from the expected result by two standard deviations, there is only about a 5% probability that the variance is due to chance.” Xerox Corp., 196 F.3d at 366. Although a result exceeding two standard deviations is sufficient to raise an inference of causation, “[t]here is no minimum statistical threshold requiring a

mandatory finding that a plaintiff has demonstrated a violation of Title VII. Courts should take a 'case-by-case approach' in judging the significance or substantiality of disparities, one that considers not only statistics but also all the surrounding facts and circumstances." Waisome v. Port Authority of New York & New Jersey, 948 F.2d 1370, 1376 (2d Cir. 1991).

Here, the plaintiff's statistical expert, Dr. Alexander Vekker, opined that each administration of the 1.5 mile run component of the PFT yielded statistically significant gender disparities in outcomes. Dr. Vekker estimated that the applicants' performance varied from a gender neutral result by more than four standard deviations in each of the three administrations of the 1.5 mile run. The defendant's sole expert on the issue of statistical disparity, Dr. Harold Brull, reviewed Dr. Vekker's statistical analysis and concluded that Dr. Vekker's calculations were accurate. Thus, the undisputed statistical evidence in the record supports the inference that the 1.5 mile run component of the PFT caused a disparate impact on female applicants.

3. Selecting the Relevant Groups for Comparison

Nevertheless, the DOC contends that the plaintiffs have "failed to prove a prima facie case because they did not choose the correct population for statistical analysis." Def.'s Opp. to Pl.'s Mot. for Summ. J. & Def.'s Mot. for Summ. J., Doc. No. 113-1, at 11. The DOC argues that the plaintiff must demonstrate that the timed 1.5 mile run had an adverse impact "on all women who took the test, not just female CO applicants." Id. at 12. In particular, the DOC seeks to pool data on the 1.5 mile run performance of individuals applying to be State Police Trooper Trainees and Protective Services Trainees. Id. at 11.

In general, the proper population for statistical analysis of disparate impact is the pool of applicants for a particular position or the eligible labor pool. See Malave v. Potter, 320 F.3d 321, 326 (2d Cir. 2003); Smith v. Xerox, 196 F.3d 358, 368 (2d Cir. 1999); Moore v. Hughes Helicopters, Inc., 708 F.2d 475, 482 (9th Cir. 1983). The eligible labor pool may be used where the employment practice at issue is likely to have dissuaded otherwise qualified applicants from applying for the position in question. See EEOC v. Joint Apprenticeship Comm., 186 F.3d 110, 119 (2d Cir. 1999) (holding that “general population data” and “potential applicant pool data” is permissible to avoid the effect of selection bias in actual applicant pool); Dothard v. Rawlinson, 433 U.S. 321, 330 (noting that the “application process might itself not adequately reflect the actual potential applicant pool, since otherwise qualified people might be discouraged from applying because of a self-recognized inability to meet the very standards challenged as being discriminatory”).⁷ The eligible labor pool may also be used in lieu of the actual applicant pool if “lack of information as to an applicant pool” would “render[] it impossible to establish a prima facie disparate impact.” Malave, 320 F.3d at 327. Here, ample data exists on the nature of the applicant pool for the CO position, and the applicant pool is large enough to determine that the gender disparity in passage rates is not attributable to chance. Therefore, the applicant pool for the CO position is the appropriate reference group for disparate impact analysis.

⁷ If the 1.5 mile run disproportionately eliminates women from consideration for the CO position, and men and women are equally aware of their ability to run 1.5 miles within the allotted time, then the administration of a 1.5 mile run test likely decreases the percentage of women in the applicant pool, and the actual disparate impact was potentially larger than the observed disparate impact. For this reason, the court would have permitted the plaintiff to introduce evidence regarding the eligible labor pool. However, the plaintiff was certainly not required to do so.

The DOC argues that, because applicants for the position of State Police Trooper Trainee and Protective Services Trainee were subject to the same 1.5 mile run, the plaintiff must establish that the 1.5 mile run imposes a disparate impact on the combined applicant pool for all three positions: CO, State Police Trooper Trainee, and Protective Services Trainee. The DOC cites Xerox for the proposition that “it is only reasonable to infer a disparate impact” from an employment practice “if all persons who were subject to the process are included in the analysis.” 196 F.3d at 370. In Xerox, the Second Circuit held that all employees at the Xerox Corporation who were subject to termination by the same “decision-making process” should be included in a determination of disparate impact. Id. at 362, 368. To establish disparate impact, the plaintiffs in Xerox had to demonstrate that, among all of the employees subject to the same “decision-making process” regarding termination, the retention rate for individuals in the “protected group” was significantly different than the retention rate for all other employees. Id. at 368.

The case at bar differs from the situation in Xerox in a critical respect. Xerox involved a pool of employees drawn from a single corporation, and the corporation used the “same decision-making process” throughout the company to determine which employees to retain. 196 F.3d at 362. By contrast, the DOC seeks to pool testing data for applicants to two other agencies, both of which conduct their hiring separately from the DOC. Although the DOC, the State Police, and the Department of Public Safety are all part of the Connecticut state government, each agency is responsible for its own hiring and fairly characterized as a separate “employer” and “respondent,” as those terms are defined in Title VII. See 42 U.S.C. § 2000e(b), (n). The DOC lacks the ability

to decide whether the State Police or the Department of Public Safety utilize a timed 1.5 mile run as part of their screening process for applicants. Indeed, in response to the instant litigation, the DOC elected in 2007 to pilot a 300 meter run in lieu of the 1.5 mile run, whereas the State Police and the Department of Public Safety have exercised their prerogative to retain the 1.5 mile run. Def.'s 56(a)(1), ¶ 38, Ex. 36, 37.

None of the cases the DOC cites in support of aggregation involved the aggregation of applicants across employers. See Stagi v. Amtrak, 391 Fed. Appx. 133, 145-147 (3d Cir. 2010) (aggregation limited to Amtrak employees); Paige v. California, 291 F.3d 1141, 1148 (9th Cir. 2002) (permitting aggregation of data on different promotional exams within the California Highway Patrol and permitting aggregation of data on all non-white officers within the California Highway Patrol); Pietras v. Board of Fire Comm'rs of Farmingville, 180 F.3d 468, 474 (2d Cir. 1999) (permitting inclusion of results of full female firefighter on physical agility test with results of probationary female firefighters within the same fire department); Eldredge v. Carpenters 46 N. California Counties Joint Apprenticeship and Training Comm., 833 F.2d 1334, 1339 (9th Cir. 1987) (aggregating admission rates to a single apprenticeship program over nine years of applications); Lilly v. Harris-Teeter Supermarket, 720 F.2d 326, 336 n.17 (4th Cir. 1983) (finding it preferable to aggregate data across years to increase statistical power and detect an ongoing "pattern or practice"); Eison v. City of Knoxville, 570 F. Supp. 11, 13 (E.D. Tenn. 1983) (pooling multiple administrations of a physical fitness test by a single police department). The court rejects the DOC's contention that aggregation is appropriate where the same screening mechanism happens to be used by multiple employers hiring for different jobs.

B. DOC's Failure to Demonstrate That Test is Job Related and Consistent with Business Necessity

Once the plaintiff meets its initial burden of establishing a prima facie violation, the burden shifts to the defendant. The defendant may rebut the plaintiff's prima facie showing by "pointing out deficiencies in data or fallacies in the [plaintiff's statistical] analysis." Gulino, 460 F.3d at 361. As discussed above, the DOC has conceded that, when the applicant pool for the CO position is viewed without reference to the applicant pool for the State Police Trooper Trainee and the Protective Services Trainee positions, a disparate impact exists.

The defendant may also overcome a plaintiff's prima facie showing of disparate impact by demonstrating that the challenged practice "is job related for the position in question and consistent with business necessity." 42 U.S.C. § 2000e-2(k)(1)(A)(i). Over the past four decades, the courts and Congress have wrangled over the meaning of the terms "job related" and "business necessity."⁸ Under one line of interpretation, a hiring practice is job related if it measures the minimum qualifications necessary to be capable of performing the job (hereafter the "Minimum Qualifications Standard"). See Lanning v. Southeastern Pennsylvania Transportation Authority, 181 F.3d 478, 481 (3d Cir. 1999) ("Lanning I") ("a discriminatory cutoff score on an entry level employment examination must be shown to measure the minimum qualifications necessary for successful performance of the job in question in order to survive a disparate impact challenge"); Green v. Town of Hamden, 73 F. Supp. 2d 192, 199 (D. Conn. 1999) ("the remaining issue is whether the passing score of 60% on this test reasonably measured

⁸ "Though the terms 'business necessity' and 'job related' appear to have semantic differences, they have been used interchangeably by the courts." Gulino v. New York State Educ. Dep't, 460 F.3d 361 (2d Cir. 2006).

the minimum entry skill level necessary for an applicant to be a competent fire fighter”). Under a second line of interpretation, a hiring practice is job related if the practice is significantly correlated with elements of work behavior that are relevant to the job (hereafter “Significantly Correlated Standard”). See Gulino v. New York State Educ. Dep’t, 460 F.3d 361, 383 (2d Cir. 2006).

The court finds the Significantly Correlated Standard governs whether a cutoff score on a physical fitness examination is job related. The court concludes that the DOC has not presented any evidence from which a reasonable jury could conclude that passing the 1.5 mile run is significantly correlated with elements of work behavior relevant to the job of a CO. Given the paucity of case law in the Second Circuit, and the lack of clear guidance from the Supreme Court, the court also considers whether the 1.5 mile run measures the minimum qualifications necessary for successful performance as a CO. Here, the court finds that the DOC is unable to present evidence that this 1.5 mile run test meets the Minimum Qualifications Standard because the DOC’s cutoff times varied by age and gender.

1. Interpretation of the Terms “Job Related” and “Business Necessity”
 - a. Case Law Prior to the Civil Rights Act of 1991

Proper interpretation of the terms “job related” and “business necessity” requires a review of the complex legal history of disparate impact litigation that followed the Supreme Court’s adoption of disparate impact liability in Griggs. 401 U.S. 424.⁹ In

⁹ In one of the leading cases assessing disparate impact in the context of a physical fitness test, the Third Circuit has carefully analyzed that history. See Lanning v. Southeastern Pennsylvania Transportation Authority, 181 F.3d 478 (3d Cir. 1999) (“Lanning I”). In Lanning I, female applicants for the position of transit police officer in the Philadelphia metropolitan rail

Griggs, the court found that requiring applicants for certain positions with Duke Power Company to obtain a high school education and pass a general intelligence test as conditions of employment operated “to disqualify [African-Americans] at a substantially higher rate than white employees,” without being “significantly related to job performance.” Id. at 425-26. In finding the employer liable, the Court wrote:

What is required by Congress is the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification.

Id. at 431. The Court continued: “The touchstone is business necessity. If an employment practice which operates to exclude [African-Americans] cannot be shown to be related to job performance, the practice is prohibited.” Id. The Court clarified 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.” Id. at 432. Finally, the court observed that the original proponents of Title VII of the Civil Rights Act of 1964 assured their critics that the Act would not affect the ability of employers to use job related tests that measured whether a “prospective applicant . . . [met] the applicable job qualifications.” Id. at 434 (emphasis omitted). Beyond this, the Court did not clarify how an employer could establish that an employment practice met the “touchstone” of “business necessity.”

system brought suit alleging disparate impact arising from a physical fitness test that required all applicants to run 1.5 miles in 12 minutes. Id. at 481-82. The Lanning I court adopted a stringent standard: “under the Civil Rights Act of 1991, a discriminatory cutoff score on an entry level employment examination must be shown to measure the minimum qualifications necessary for successful performance of the job in question in order to survive a disparate impact challenge.” Id. at 481. The abbreviated historical overview that follows borrows from the Third Circuit’s helpful timeline in Lanning I. The applicable line of cases has also been usefully summarized through the year 2000 in David E. Hollar, Comment, Physical Ability Tests and Title VII, 67 U. Chi. L. Rev. 777 (2000).

In Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975), the Court purported to directly answer the question: “What must an employer show to establish that pre-employment tests . . . discriminatory in effect, though not in intent, are sufficiently ‘job-related’ to survive challenge under Title VII?” Id. at 408. The Court embraced the EEOC Guidelines which require that tests be “shown, by professionally acceptable methods, to be ‘predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.’” Id. at 431 (quoting 29 C.F.R. § 1607.4(C)). Under this standard, the Court concluded that the Albemarle Paper Company had failed to properly validate two tests it used to decide whether an individual was promoted to a more skilled line of work. Id. 410-11.

The Court next addressed the issue of business necessity in Dothard v. Rawlinson, 433 U.S. 321 (1977). See Lanning I, 181 F.3d at 486. In Dothard, the Court rejected the application of minimum height and weight requirements for prison guards because they caused a disparate impact on women that was not justified by job-relatedness. Id. at 331-32. The court observed that the employer “produced no evidence correlating the height and weight requirements with the requisite amount of strength thought essential to good job performance. Indeed, they failed to offer evidence of any kind in specific justification of the . . . standards.” Id. at 331. Moreover, if a certain level of strength was required to fulfill the tasks of a prison guard, that strength could be measured directly. Id. at 332. In general, the Dothard decision was consistent with the Significant Correlation Standard articulated in Albemarle. In a footnote, the Court included language implicating the Minimum Qualifications Standard,

explaining that an employer could establish business necessity by showing that the practice was “necessary to safe and efficient job performance.” Id. at 331, n.14 (emphasis added).

A turning point in the “business necessity” jurisprudence came in 1989, with the Court’s decision Wards Cove Packing Company, Inc. v. Atonio, 490 U.S. 642 (1989). The Court held that the plaintiffs had not made out a prima facie case of disparate impact. Because the Court remanded the case for further proceedings, the Court elected to provide guidance on the issue of business necessity. Id. at 655, 658. The Court refined the concept of business necessity, observing that “there is no requirement that the challenged practice be ‘essential’ or indispensable’ to the employer’s business for it to pass muster: this degree of scrutiny would be almost impossible for most employers to meet, and would result in a host of evils.” Id. at 659. Rather, “the dispositive issue is whether a challenged practice serves, in a significant way, the legitimate employment goals of the employer.”¹⁰ Id. The language in Wards Cove can only be understood to reject the Minimum Qualifications Standard as too onerous for employers.

b. The Civil Rights Act of 1991

Congress responded to the Wards Cove decision by attempting to codify the disparate impact concept as it had been articulated in the Court’s jurisprudence prior to Wards Cove. See Civil Rights Act of 1991, Pub. L. No. 102-166, § 3 (codified at 42 U.S.C. § 1981 (note)). Congress passed the Civil Rights Act of 1991, under which an

¹⁰ The Court in Wards Cove also modified the burdens of proof in disparate impact claims, stating that the employer bore only the burden of production at the business necessity stage, with the burden of persuasion remaining with the plaintiff. Id. at 659; see also Lanning I, 181 F.3d at 487.

employer may only overcome a prima facie showing of disparate impact by “demonstrat[ing] that [a] challenged practice is job related for the position in question and consistent with business necessity.” 42 U.S.C. § 2000e-2(k)(1)(A) (emphasis added). Section 3 of the Act expressly declares that one of its purposes was “to codify the concepts of ‘business necessity’ and ‘job related’ enunciated by the Supreme Court in Griggs . . . and in other Supreme Court decisions prior to Wards Cove.” Pub. L. No. 102-166, § 3.

The Act also designated a three paragraph “interpretive memorandum” as the exclusive “legislative history” of the Act when “construing or applying, any provision of this Act that relates to Wards Cove – Business necessity / cumulation / alternative business practice.” Pub. L. No. 102-166, § 105(b) (codified as 42 U.S.C. § 1981 (note)). That interpretive memorandum stated, in pertinent part: “The terms ‘business necessity’ and ‘job related’ are intended to reflect the concepts enunciated by the Supreme Court in Griggs v. Duke Power Co., 401 U.S. 424 (1971), and in other Supreme Court decisions prior to Wards Cove Packing Co. v. Atonio, 490 U.S. 642 (1989).” 137 Cong. Rec. S15273-01 (Oct. 25, 1991). Section 3 of the statute, along with the designated interpretive memorandum, make clear that Congress viewed the “business necessity” language in Wards Cove as a departure from the earlier “business necessity” doctrine articulated in the Griggs line of cases. Because Wards Cove rejected any “requirement that the challenged [hiring] practice be ‘essential’ or indispensable’ to the employer’s business,” 490 U.S. at 659, it follows that Congress sought to require hiring practices that cause disparate impact to measure the minimum qualifications necessary to be capable of performing the job in question. See Lanning I, 181 F.3d at 481.

On the other hand, the Civil Rights Act of 1991 was adopted to restore the Supreme Court's pre-Wards Cove jurisprudence, but language embracing the Minimum Qualification Standard is generally absent from the cases leading up to Wards Cove. The statutory history of the Civil Rights Act of 1991 also points toward an interpretation of the phrase "consistent with business necessity" as requiring the Significant Correlation Standard. In his dissenting opinion in Lanning I, Judge Weis recounts:

Some members of Congress were displeased with the results in Wards Cove and argued for a stricter standard of business justification based on their reading of pre-Wards Cove cases. After two years of legislative struggle, Congress and the President agreed upon a compromise bill. . . .

The 1990 bill, which had been vetoed by the President, had used the phrase "*required by* business necessity," rather than "*consistent with* business necessity," as used in the 1991 Act. The substitution of the word consistent was considered to indicate a standard less stringent than would "required."

Lanning I (Weis, J., dissenting), 181 F.3d at 496-97. Viewed through the lens of statutory history, the selection of the phrase "consistent with business necessity" rather than "required by business necessity" implies that Congress and the President codified the "business necessity" concept such that a hiring practice would be characterized as "job related" if that hiring practice was "significantly correlated with elements of work behavior that were relevant to the job." See Albemarle Paper Co., 422 U.S. at 431; 29 C.F.R. § 1607.4(C).

c. Case Law Following the Civil Rights Act of 1991

Subsequent case law points toward this less stringent standard for establishing that a practice is consistent with business necessity. The Supreme Court indirectly addressed the issue of "business necessity" in the context of Title VII in Ricci v.

DeStefano, 129 S. Ct. 2658 (2009).¹¹ In Ricci, the Court decided whether the City of New Haven could refuse to certify promotional testing results for the New Haven Fire Department on the grounds that certifying the results could expose the City to liability for disparate impact. Id. at 2664-65. The Court held that the City of New Haven lacked a “substantial basis in evidence” that certifying the results would expose the City to disparate impact liability. Id. at 2676, 2678. In Ricci, the parties agreed that a prima facie case of disparate impact could be established. Id. at 2677. However, the Court concluded that the prima facie case would be overcome because the tests were job related and consistent with business necessity. Id. at 2678 (“There is no genuine dispute that the examinations were job related and consistent with business necessity. The City’s assertions to the contrary are ‘blatantly contradicted by the record’”) (citation omitted). Although the Court did not discuss its view of the proper interpretation of “business necessity,” the Court appears to have reached its conclusion on the basis of evidence in the record that questions on the firefighter’s written and oral exams were “relevant” to the duties of the positions in question. Id. at 2678-79.

In cases following the Civil Rights Act of 1991, the Second Circuit has generally embraced the Significantly Correlated Standard for business necessity that the Supreme Court implicitly endorsed in Ricci. See Gulino v. New York State Educ. Dep’t, 460 F.3d 361, 383 (2d Cir. 2006). In Gulino, the Court of Appeals vacated a district court’s decision that a general knowledge test component of New York State’s public

¹¹ Other than the decision in Ricci, the Supreme Court has been virtually silent since the Civil Rights Act of 1991 as to the interpretation of the term “consistent with business necessity.” The Court briefly referenced the term “business necessity” in Smith v. City of Jackson, Miss., 544 U.S. 242, 243 (2005) (“Unlike the business necessity test, which asks whether there are other ways for the employer to achieve its goals that do not result in a disparate impact on a protected class, the reasonableness inquiry [under the ADEA] includes no such requirement.”).

school teacher certification program was job related. Id. at 383-88. The Gulino court adopted the Significantly Correlated Standard first articulated in Albemarle Paper Co., stating that “discriminatory tests are impermissible unless shown, by professionally acceptable methods, to be predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.” 460 F.3d at 383 (quoting Albemarle Paper, 422 U.S. at 431). The court then found that the district court below had articulated the wrong legal standard for validating the employment test. The court also emphasized that Guardians Assoc. of NYC Police Dep’t, Inc. v. Civil Service Comm’n of City of NY, 630 F.2d 79 (2d Cir. 1980), “is still the law in this Circuit,” Gulino, 460 F.3d at 385.

The Guardians decision, however, could reasonably be interpreted as endorsing the Minimum Qualifications Standard. In Guardians, the Second Circuit held that an entrance exam for the New York City Police Academy violated Title VII’s disparate impact provisions. The New York City Police Department required a pool of 12,000 eligible candidates, so the Department selected the passing score for the entrance exam by “simply set[ting] the passing score at the score achieved by the 12,000th highest scoring candidate.” Id. at 85. The court stated that “[a] cutoff score unrelated to job performance may well lead to the rejection of applicants who were fully capable of performing the job. When a cutoff score unrelated to job performance produces disparate racial results, Title VII is violated.” Id. at 105. Here, the Guardians court seemed concerned that a cutoff score set higher than absolutely necessary could exclude otherwise qualified applicants. The court also endorsed the EEOC’s recommendation that cutoff scores “normally be set so as to be reasonable and

consistent with normal expectations of acceptable proficiency within the work force.” Id. (quoting 29 CFR § 1607.5(H)). The term “acceptable proficiency” connotes a set of minimum qualifications for a position. Finally, the Guardians court urged that there be “some independent basis for choosing the cutoff” by “using a professional estimate of the requisite ability levels, or, at the very least, by analyzing the test results to locate a logical ‘break-point’ in the distribution of scores.” Id. at 105. Again, reference to “requisite ability levels” implies the application of the Minimum Qualifications Standard. This interpretation is perhaps undermined, however, by the court’s willingness to forgo a “professional estimate of the requisite ability levels” and proceed with a cutoff score set to a natural “break-point” in the distribution of scores. See id.

The Second Circuit has only tangentially addressed the selection of cutoff scores for tests of physical ability. See Pietras v. Board of Fire Commissioners of the Farmingville Fire District, 180 F.3d 468, 472 n. 5 (2d Cir.1999). In Pietras, a town fire department selected a four minute cutoff time for the completion of a physical agility test for probationary firefighters. Id. at 471. That cutoff time was selected by “taking the average of all the test scores and then arbitrarily adding some extra time.” Id. at 472, n.5. The court observed that the defendant bore the burden of showing that the test was job related, but the defendant had presented “no evidence at all to indicate that the time chosen for the test reflected the needs of the job.” Id. Thus, the Second Circuit found that – even assuming the town was challenging the district court’s finding that the cutoff score was not job related – the challenge was “without merit.” Id.

The dicta in Pietras does not directly answer whether a cutoff score for a physical exam must be set to measure the minimum qualifications necessary for the position in

question or whether passing an exam with a particular cutoff score must merely be significantly correlated with an applicant's performance as an employee. The language requiring the chosen cutoff time to "reflect[] the needs of the job" points toward a Minimum Qualifications Standard, but the discussion in Pietras is too brief and too ancillary to the decision to glean the requisite degree of specificity.

The Pietras case is factually analogous to the instant matter in that it involved a test of physical ability. The Pietras case, however, can also be distinguished. The physical test in Pietras involved simulations of activities a firefighter might actually be required to perform, whereas the parties in this case agree that the 1.5 mile run test was only meant to measure aerobic capacity. Aerobic capacity is a prerequisite to a number of activities a CO might have to perform. Running itself is not a close approximation to the typical duties of a CO. The Second Circuit's discussion of business necessity in Pietras is also too oblique to be usefully applied to the present circumstances.

The court concludes that the Significantly Correlated Standard articulated by the Second Circuit in Gulino applies to this case. 460 F.3d at 383. The statutory text of the Civil Rights Act of 1991 does not define the terms "job related" and "consistent with business necessity." The statutory text and the legislative history of the Civil Rights Act of 1991 both indicate that the Act was intended to restore the meaning of these terms to their status prior to Wards Cove. See Civil Rights Act of 1991, Pub. L. No. 102-166, § 3; Interpretive Memorandum, 137 Cong. Rec. S15273-01 (Oct. 25, 1991). Wards Cove clearly prohibited a Minimum Qualifications Standard. 490 U.S. at 659. However, a fair reading of Griggs, Albemarle Paper Co., and Dothard suggests that the Supreme Court had never embraced a Minimum Qualifications Standard. Indeed, Albemarle Paper Co.

explicitly introduced the Significant Correlation Standard. 422 U.S. at 431. In Gulino, the Second Circuit properly discerned that “the basic rule has always been that ‘discriminatory tests are impermissible unless shown, by professionally acceptable methods, to be predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job.’” 460 F.3d at 383. For that reason, the court applies the Significant Correlation Standard to this case.

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. Under this interpretation, any test used to select employees must first be “valid,” as that term is defined in the EEOC Guidelines. For example, in the context of a criterion-related validity study, “[e]vidence of the validity of a test or other selection procedure . . . should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance.” 29 C.F.R. § 1607.5(B). Second, if the selection test is valid, and the employer refuses to hire any applicant scoring below a certain cutoff, then that cutoff score must measure the minimum qualifications necessary for success in the position.¹² 29 C.F.R. § 1607.5(H) (“Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force.”) Thus, better performance on the test must be significantly correlated with better performance as an employee, and any cutoff score must measure the minimum qualifications necessary to competently perform the job.

¹² Employers do not always impose cutoff scores for their selection tests. An employer might use test performance as one of several factors in an overall hiring decision, rather than barring all applicants scoring below a certain threshold.

For example, in Albemarle Paper Co., the Supreme Court focused on whether two “general ability tests” administered by an employer were “predictive of or significantly correlated with important elements of work behavior.” 422 U.S. at 427, 431 (citing 29 C.F.R. § 1607.5(B)). The Court observed that the validation study undertaken by the Albemarle Paper Co. showed significant correlation for less than half of the lines of employment and was unreliable due to an “odd patchwork of results” and a performance measurement that was “extremely vague and fatally open to divergent interpretations.” *Id.* at 431-33. The Court concluded that “the District Court erred in concluding that Albemarle had proved the job relatedness of its testing program.” *Id.* at 435-36. Because the Court concluded that the selection mechanism itself was invalid, the Court never addressed whether the cutoff scores on these tests were appropriate.

Contrast this with the Lanning cases, where running significant distances constituted an essential job task of a transit police officer. See Lanning II, 308 F.3d 286, 290 (3d Cir. 2002) (“a SEPTA transit police officer must be ready and able to apprehend not just the numerous sedentary, petty criminals, but also the fleet-footed few who, from time to time, wreak serious harm on the people of Philadelphia”). An applicant’s performance on a 1.5 mile timed run was undoubtedly correlated with whether that applicant could later run a similar distance in the course of her potential work as a transit police officer. Although the Third Circuit articulated a Minimum Qualifications standard, it did so in the context of a cutoff score: the requirement that all applicants successful run 1.5 miles in less than 12 minutes. Lanning I, 181 F.3d 478, 481, 487 (3d Cir. 1999) (“a discriminatory cutoff score . . . must be shown to measure

the minimum qualifications necessary for successful performance,” and “a discriminatory cutoff score is impermissible”).

Fortunately, harmonization of the case law is not required for the proper disposition of these Motions. The 1.5 mile run times imposed by DOC during the 2004 and 2006 physical fitness examinations cannot be characterized as “job related for the position in question and consistent with business necessity” under the Significant Correlation Standard or the Minimum Qualification Standard.

2. Whether the DOC Has Shown That the Cut Times in the 1.5 Mile Run Represent a Minimum Qualification Necessary to Perform the CO Position

Under the Third Circuit’s interpretation in Lanning I, a “discriminatory cutoff score is impermissible unless shown to measure the minimum qualifications necessary for successful performance of the job in question.” 181 F.3d at 481. The parties agree that the 1.5 mile run is a test that measures an individual’s aerobic capacity. Def.’s 56(a)(1), ¶ 12; Pl.’s 56(a)(2), ¶ 12. The DOC cannot plausibly argue that a time of 12:25 for 21-29 year-old men is a valid predictor of the aerobic capacity minimally necessary for successful completion of the tasks of a Correction Officer, if the DOC also permitted 21-29 year-old women to complete the 1.5 mile run in 14:49, and 50 year-old women to complete the 1.5 mile run in 17:14. By definition, cutoff times that vary by gender and age cannot represent a measure of the minimum aerobic capacity necessary for successful performance as a CO. Only a single cutoff time could meet this standard.¹³

¹³ If the DOC had used a single cutoff time for the 1.5 mile run, such a showing might have been possible. For example, in Lanning v. Southeastern Pennsylvania Transportation Authority, 308 F.3d 286 (3d Cir. 2002) (“Lanning II”), the Third Circuit credited a study that “found that individuals who passed the run test had a success rate on” critical policing tasks that ranged from 70% to 90%, whereas individuals who failed the run test had a success rate between 5% and 20%. Id. at 291. “The District Court found that such a low rate of success was

The defendant purports to rely on the Cooper Institute's labeling of performance at the 40th percentile as "fair" rather than "poor." See Def.'s 56(a)(1), ¶¶ 10, 13. However, the Cooper Institute specifically warns its consumers not to use percentile rankings as hiring standards. Pl.'s 56(a)(1), ¶ 29, Ex. 20, at 3. The Cooper Institute states that "percentile scores (whether age and gender norms or single norms) have no validity data for predicting who can and who cannot do the job." Pl.'s 56(a)(1), ¶ 30, Ex. 21, at 9.¹⁴ Moreover, the EEOC Uniform Guidelines on Employee Selection Procedures "specifically rule[s] out" assumptions of validity based on "descriptive labels" – here, the arbitrary designation of the 40th percentile as "fair" by the Cooper Institute. See 29 C.F.R. § 1607.9.

unacceptable for employees who are regularly called upon to protect the public. In so doing, the District Court implicitly defined 'minimum qualifications necessary' as meaning 'likely to be able to do the job.'" Id.

¹⁴ The Cooper Institute also warns against using percentile rankings for the purpose of meeting the Significantly Correlated Standard. The Cooper Institute states that "[u]sing percentile rankings of the Cooper norms for standards is not defensible" because the "percentile rankings do not predict the ability to do the job and do not demonstrate criterion validity." Pl.'s 56(a)(1), ¶ 29, Ex. 20, at 3. Criterion validity is demonstrated by identifying criteria that indicate successful job performance and then correlating test scores with those identified criteria. Guardians, 633 F.2d at 242, n.16.

3. Whether the DOC Has Shown That the Cut Times in the 1.5 Mile Run Are Significantly Correlated with Elements of Work Behavior Relevant to the Job.

Under the Second Circuit's decision in Gulino, "discriminatory tests are impermissible unless shown, by professionally acceptable methods, to be predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated." 460 F.3d at 383 (citation omitted). Here, the DOC "produced no evidence correlating" linking successful completion of the timed 1.5 mile run test "with the requisite amount of" aerobic capacity "thought essential to good job performance. Indeed, they failed to offer evidence of any kind in specific justification of the . . . standards." Dothard, 433 U.S. at 331.

All three of the defendant's experts on the issue of business necessity admit that they have not empirically demonstrated that a CO applicant's passage of the 1.5 mile run is correlated with that applicant's performance on particular job tasks as a CO. Dr. Brull testified that he had not observed any statistically significant correlation between aerobic capacity and CO performance. Pl.'s 56(a)(1), ¶ 23, Ex. 19, at 125. Dr. Brull stated that "it's reasonable to assume [that] higher levels of fitness correlate with more positive outcomes and fewer negative outcomes," but did not provide any evidence linking successful completion of the 1.5 mile run with objective measurements of performance as a CO. See Pl.'s 56(a)(1), ¶ 24, Ex. 19, at 116.

The DOC bears the burden of establishing business necessity. DOC has not made any attempt to demonstrate that passing the 1.5 mile run is "predictive of or significantly correlated with important elements of work behavior which comprise or are

relevant to the job or jobs for which candidates are being evaluated.” Albemarle Paper Co., 422 U.S. at 431 (quoting 29 C.F.R. § 1607.4(C)). On the record before the court, no reasonable jury could conclude that the cut times used by the DOC for the 1.5 mile run were job related for the position in question and consistent with business necessity.

C. Alternative Employment Practices

If the DOC had met its burden “of showing that the challenged practice is job related, the plaintiff [could] only prevail by showing that ‘other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in efficient and trustworthy workmanship,’” Gulino, 460 F.3d at 382, and the DOC “refuses to adopt such alternative employment practice.” 42 U.S.C. § 2000e-2(k)(1)(A)(ii). Because the DOC has not shown the 1.5 mile run to be job related and consistent with business necessity, the court does not reach the issue of whether the DOC refused to adopt an alternative employment practice that would serve the employer’s legitimate needs.

V. **CONCLUSION**

For the foregoing reasons, the Plaintiff’s Motion for Summary Judgment [**Doc. No. 106**] is **GRANTED**, and the Defendant’s Motion for Summary Judgment [**Doc. No. 113**] is **DENIED**. The court will schedule a status conference to discuss the damages phase of this case.

SO ORDERED.

Dated at Bridgeport, Connecticut this 5th day of May, 2011.

/s/ Janet C. Hall
Janet C. Hall
United States District Judge