Arndt et al v. City of Colorado Springs Doc. 156

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLORADO
Senior District Judge Richard P. Matsch

Civil Action No. 15-cv-00922

REBECCA ARNDT,
NICOLE BALDWIN,
CATHY BUCKLEY,
STACEY CLARK,
DONYA DAVIS,
JULIE GARRETT,
CAROLYN GRAVES,
SAMANTHA LEMBERGS,
JENNIFER LEWIS,
GERALDINE PRING,
MAGDALENA SANTOS, AND
TERRY THRUMSTON,

Plaintiffs,

v.

CITY OF COLORADO SPRINGS,

Defendant.

FINDINGS, CONCLUSIONS AND ORDER DECIDING
PLAINTIFFS' CLAIM OF DISPARATE IMPACT DISCRIMINATION

The Second Amended Complaint, filed February 1, 2016, includes a claim that the use of a physical fitness test to determine continuation of employment as Colorado Springs Police Officers has had a disparate impact on women officers over 40 years of age in violation of Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000e et seq. That being an equitable claim to be determined by the Court, a motion to bifurcate it from the other claims was filed on September 2, 2016 (doc. 102). After hearing the defendant's opposing arguments, the

Dockets.Justia.com

Court denied that motion, finding that the factual questions were too common with the jury claims to be determined at an earlier bench trial.

The parties then filed motions in limine and under Fed. R. Evid. 702, challenging the opinions in the reports of John Peters; Dan Montgomery; Arthur Weltman, Ph.D.; Kurt Kraiger, Ph.D.; and Norman D. Henderson, Ph.D., submitted under Fed. R. Civ. P. 26(a)(4)(B) and their deposition testimony. After reviewing the papers filed on those motions, the Court determined to proceed with the bench trial permitting the witnesses to testify and considering the objections in determining the credibility of those witnesses. There was no objection and the trial proceeded after denial of the motions in open court on October 31, 2016.

After consideration of the evidence submitted at trial and the written and oral arguments of counsel the Court now makes the findings of fact and conclusions of law required by Rule 52 in the following narrative form.

In 2009, Chief of Police Richard Myers decided to implement physical fitness testing for all officers working in the Colorado Springs Police Department ("Department").

The City of Colorado Springs contracted with Human Performance Systems, Inc. ("HPS"), a company based in Beltsville, Maryland, to develop a physical abilities test for use by the Department to evaluate all of its officers for fitness for duty. The policy determination was that all officers must demonstrate the ability to perform all of the tasks of a patrol officer and if an officer failed the result could be termination of employment.

On the recommendation of HPS, the Department adopted a four-part physical abilities test ("PAT"), comprised of a one-minute sit-up test; a one-minute push-up test; an agility run; and a running test known as a BEEP test. Tr. Vol. VI (Eells) at 556:8 - 557:14; Ex. 2

("Validation Report") at CSPD-PAT 00434. These four tests were selected because they were considered to be a significant predictor of job performance and met the Department's administrative decision to conduct the testing indoors. *Id.*

The scoring system adopted was a compensatory scoring method. With that method, a participant's scores on each component skill test are combined into one final score and there is only one overall cut-off score. For the PAT, a maximum of 8 points was assigned to each of the four skills tests, for a total maximum score of 32 points.[1] The passing score was set at twenty points, with at least one point on each of the four components. Validation Report at CSPD-PAT 00461 - 68. The same passing score applied to male and female officers.

In the early months of 2013, the Department administered the PAT to applicants. A total of 421 recruits took the PAT (343 males and 78 females). Of those, 50% of the females failed, compared to a 6% of the males. Henderson Ex. 4.

In 2013, the Department administered a practice test of the PAT to all incumbent officers. Tr. Vol. VI (Eells) at 571:21 - 572:24. That practice test was given to assist officers in assessing their physical fitness in preparation for mandatory testing. Another objective was to determine whether the test had an adverse impact on any particular group of officers. *Id.*

Results of the 2013 practice test showed that 421 of 467 men passed, for a passing rate of 90.5%. Forty (40) of 67 women passed the practice test, for a passing rate of 59.7%. Kraiger Ex. 10, ¶ 5 at p. 3. Officers who failed the 2013 practice PAT were not disciplined or subjected to any adverse employment action.

---

[1]Table 41 of the Validation Report is a scoring table which shows the point values attributed to various performance levels on the four test components. Validation Report at CSPD-PAT 00467.

On September 3, 2014, the Department – then under the direction of Police Chief Peter Carey – issued General Order 1915, stating that all sworn police officers employed by the Department were required to participate in an annual physical fitness test consisting of the push-up test; Illinois agility run; sit-up test, and the BEEP test. That order announced that "any employee who does not meet the Minimum Performance Standard will be placed on light duty and on a Performance Improvement Plan (PIP) until he/she can successfully complete the process with a minimum score of twenty (20)." Montgomery Ex. 6. The order stated that officers who failed the test could retake it at least once per month (or more frequently) and were required to pass within six months. The order stated that officers placed on light duty as a result of unsatisfactory PAT performance were prohibited from participating in any promotional or specialized selection process and that failure to pass within the six-month period could result in termination of employment for failure to meet the minimum qualifications of a Colorado Springs police officer. *Id.*

The Department issued Bulletin 548-14 on December 14, 2014, stating *inter alia,* that officers placed on light duty due to unsatisfactory PAT performance were prohibited from responding to a scene or any type of field work environment; were subject to restrictions with respect to overtime work; could not be placed on-call or standby or have a take-home vehicle; were not allowed to be in uniform or wear any attire that would identify him/her as a police officer, and were subject to certain restrictions with respect to the carrying of a firearm. Montgomery Ex. 7.

At the conclusion of the 2014 testing cycle, approximately 96 % of all officers passed the PAT on their first attempt, and the majority of those who initially failed ultimately passed on

subsequent attempts. Tr. Vol. VI (Eells) at 578:14 - 579:17. Of those who never passed the PAT, some left the Department and some did not retake the test due to injuries.

All twelve plaintiffs initially failed the PAT. Nine of them passed on subsequent testing. Sergeant Garrett, Detective Thrumston and Lieutenant Santos have not passed the test.

The 2014 PAT was the only complete mandatory testing cycle. According to the parties' stipulation and entry of preliminary injunction in this action, testing has been halted. The Department has not terminated the employment of any officer for failure to pass the PAT.

"Title VII forbids ... 'practices that are fair in form, but discriminatory in operation,' most often referred to as 'disparate impact' discrimination." *Tabor v. Hilti, Inc.*, 703 F.3d 1206, 1220 (10th Cir. 2013) (quoting *Lewis v. City of Chicago*, 560 U.S. 205 (2010)); 42 U.S.C. § 2000e–2(k). "The disparate impact 'doctrine seeks the removal of employment obstacles, not required by business necessity, which create built-in headwinds and freeze out protected groups from job opportunities and advancement.'" *Tabor*, 703 F.3d at 1220 (quoting *E.E.O.C. v. Joe's Stone Crab, Inc.*, 220 F.3d 1263, 1274 (11th Cir. 2000)).

A plaintiff claiming disparate impact discrimination must establish that an identifiable employment practice or policy causes a significant disparate impact on a protected group. 42 U.S.C. § 2000e-2(k)(1)(A)(i); *Tabor*, 703 F.3d at 1220. If the plaintiff makes that showing, the burden shifts to the employer "to demonstrate 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); *Tabor*, 703 F.3d at 1220-21. If the employer demonstrates business necessity, the plaintiff may still prevail 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." *Tabor*, 703 F.3d at 1221 (quoting *Ricci v. DeStefano*, 557 U.S. 557, 578 (2009)); 42 U.S.C. § 2000e–2(k)(1)(A)(ii).

Plaintiffs have identified a specific employment practice. They challenge the Department's employment policies set forth in General Order 1915 and Guidance 548-14. Plaintiffs complain that requiring all sworn officers to pass the PAT annually or risk disciplinary actions, including termination of employment, has a disparate impact on female officers over the age of 40 and/or all female officers.

Did the plaintiffs prove by a preponderance of the evidence that the use of the PAT in 2014 had a discriminatory impact on women police officers exposing them to termination of their employment after many years of satisfactory performance? The plaintiffs have used statistical evidence to demonstrate that effect.

"Statistical evidence is an acceptable, and common, means of proving disparate impact." *Tabor*, 703 F.3d at 1222 (quoting *Carpenter v. Boeing Co.*, 456 F.3d 1183, 1196 (10th Cir. 2006)).

The Equal Employment Opportunity Commission ("EEOC") has issued a guideline, known as the "four-fifths" rule, which states that a disparity of 20% will be considered evidence of adverse impact. The EEOC's "four-fifths" rule provides in part:

> 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 by Federal enforcement agencies 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 or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group. . . .

29 C.F.R. § 1607.4(D).

According to the EEOC, application of the four-fifths rule involves the following four steps:

> (1) calculate the rate of selection for each group (divide the number of persons selected from a group by the number of applicants from that group).
>
> (2) observe which group has the highest selection rate.
>
> (3) calculate the impact ratios, by comparing the selection rate for each group with that of the highest group (divide the selection rate for a group by the selection rate for the highest group).
>
> (4) observe whether the selection rate for any group is substantially less (i.e., usually less than 4/5ths or 80%) than the selection rate for the highest group. If it is adverse impact is indicated in most circumstances.

EEOC, Adoption of Questions and Answers To Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures, Question 12, 44 Fed. Reg. 11996 (March 2, 1979).[2]

The United States Court of Appeals for the Tenth Circuit has recognized that the EEOC's guideline is persuasive, although not controlling on the courts. *Tabor*, 703 F.3d at 1222.

A plaintiff claiming disparate impact must show that a disparity is statistically significant. That requires evidence addressing "the likelihood that the disparity between groups is random, i.e., solely the result of chance. [Statistical significance] is expressed in terms of standard errors or standard deviations." *Tabor*, 703 F.3d at 1223. "The Supreme Court has recognized that a disparity of more than two or three standard deviations in a large sample makes 'suspect' the contention that the differential occurs randomly." *Id.* (quoting *Carpenter*, 456 F.3d at 1195; *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 308 n. 14 (1977)).

---

[2]Found at https://www.eeoc.gov/policy/docs/qanda_clarify_procedures.html.

Plaintiffs rely primarily on the reports and testimony of Dr. Kurt Kraiger, an industrial-organizational psychologist. Dr. Kraiger is qualified by education and experience to conduct statistical analyses of employment testing.

Using the data provided by the Department resulting from both the 2013 pre-test and the 2014 mandatory test, Dr. Kraiger presented six sets of statistical analysis, comparing different groupings of officers who took the PAT in 2013 and 2014. Kraiger Ex. 10. Dr. Kraiger used the EEOC's four-fifths rule to assess disparities in the passing rates of the groups he compared. To assess the statistical significance of disparities, Dr. Kraiger used the "chi-square (χ^2)" test. The chi-square test is an accepted statistical method of assessing the probability that a disparity is due to chance. *See Powers v. Ala. Dep't of Educ.*, 854 F.2d 1285, 1298 (11th Cir. 1988) ("Several courts have approved chi-square analysis as an alternative to standard deviation analysis . . . .").

Not all of Dr. Kraiger's comparisons are relevant. In one comparison, he combined the results of the 2013 practice test and the 2014 mandatory test to demonstrate an adverse impact on the women officers. As Dr. Norman Henderson testified that is inappropriate because the same people took the tests, resulting in duplication.

In their pleadings the plaintiffs asserted that women over 40 should be considered a potential class adversely impacted by the physical test requirement. Age discrimination is also alleged in a separate claim for violation of the ADEA and such discrimination may also be the subject of a disparate impact theory of recovery. *Smith v. City of Jackson, Mississippi*, 544 U.S. 228 (2005). The two may not be conflated because the defenses to liability are different. The scope of liability under the ADEA is more narrow.

At oral argument, plaintiffs' counsel conceded that for this disparate impact claim the protected class is incumbent women police officers. While the ADEA and Title VII claims may not be mixed together, age is not irrelevant to this case. There is disparity in the test results of men and women over forty tested in 2014. 29 of 43 women passed (67.4%). 317 of 326 men passed (97.2%). The ratio of pass rates is 69.36%.

There was a difference in the results of women over 40 and younger women. These differences suggest that there may be a difference in the effects of the natural aging process between men and women but the record does not include evidence adequately supporting that specific finding.[3]

Evidence presented at trial supports the conclusion that the PAT disparately impacts women. As set forth above, when the Department used the PAT to screen applicants in 2013, 50% of the females failed, compared to a 6% of the males. Henderson Ex. 4. Those test results are significant and are evidence of adverse impact on women. If the plaintiffs were applicants for employment as patrol officers it would be clear that the use of the PAT as a screening method presented a gender barrier to this employment.

The results of the 2013 practice test administered to all officers are also significant. Dr. Kraiger applied the four-fifths rule to those results and opined that the ratio of men's passing rate and women's passing rate showed an adverse impact on the group of all women. Kraiger Ex. 10, p.3, ¶ 5; Tr. Vol. II at 179:3-13; 181:6 -182:25; 184:21 to 186:7. The City does not dispute Dr. Kraiger's statistical analysis of those results. The City instead contends that the

[3]Dr. Norman Henderson, an expert witness retained by the City, made the following observation in his report: "In the present case, the performance of older females is a simple additive of sex differences and age differences." Henderson Ex. 2 at p. 22.

results of the 2013 practice test should be given no weight, suggesting that officers lacked motivation to prepare and make their best effort for that test. That argument is not persuasive because if insufficient motivation was a factor, that would have been so for both men and women.

In 2014, when the mandatory test was administered, 555 men were tested, and 544 achieved a passing score on their first attempt, for a passing rate of 98%. Kraiger Ex. 10, ¶ 4 at pp. 2 - 3. Seventy-nine (79) women took the 2014 test, and 64 achieved a passing score on their first attempt, for a passing rate of 81%. *Id.* It is undisputed that for that test, the ratio of passing rates for all women and all men is 82.6%. That impact ratio is slightly above the 80% threshold.

Dr. Kraiger applied chi-square analysis and opined that the disparity between men and women in the 2014 test has statistical significance. Kraiger Ex. 10 at p. 3. He stated that if men and women passed at similar rates, the disparity in the women's and men's 2014 passing rates would occur less than 1 percent of the time. *Id.* The defendant contends that Dr. Kraiger's chi-square analysis should be disregarded, arguing that than an assessment of statistical significance is unnecessary and improper when the four-fifths threshold is not met.

This case does not completely depend on Dr. Kraiger's opinion. The EEOC guideline expressly notes that smaller differences in the selection rate may nevertheless constitute adverse impact, where they are significant in practical terms, discouraging applicants disproportionately on grounds of sex.

Notably, the guideline is designed for screening applicants for employment. The Department's use of the PAT for termination of employment is without precedent. Discouraging women from applying for employment is quite different from the fear of losing the job you have

had for years because you can't do these exercises according to the scoring system used. The plaintiffs are incumbent officers who have been performing their duties satisfactorily as shown by their evaluation reports.

The Department contends that their claim of disparate impact is defeated by the fact that most of them passed in later attempts. That ignores the evidence of plaintiffs Cathy Buckley, who passed on the fifth attempt, and Carolyn Graves, who passed on the second try. They hired physical trainers and spent many hours in training to achieve a passing score.

The failure to pass initially has had a devastating effect on the plaintiffs who have had to endure the indignity of being denied recognition as a police officer by the restrictions imposed by the Department. They have been shamed and ostracized.

The plaintiffs' evidence of disparate impact is sufficient to consider whether the defendant's defense of business justification has been proved. This requires consideration of the validity of the PAT.

To avoid liability for the discriminatory impact of the PAT requirement for all officers, the City must prove by a preponderance of the evidence that this requirement is job-related and consistent with business necessity.

"A business justification proffered by an employer 'must have a manifest relationship to the employment in question.'" *Faulkner v. Super Valu Stores*, 3 F.3d 1419, 1429 (10th Cir. 1993) (quoting *Griggs v. Duke Power Co.*, 401 U.S. 424, 432 (1971)).

The City asserts that the Department's PAT policy was implemented to promote a culture of fitness in the Department. The City further asserts that the PAT policy – by promoting and ensuring officers' physical fitness – enhances public safety and reduces the risk of on-the-job

injuries. The City contends that it is in the best position to determine which policies are necessary for its business and because the field of employment involves public safety, the Court should defer to the City's determinations about job-relatedness and business necessity.

Promoting physical fitness in the Department's employees is a laudable goal, particularly for employees tasked with protecting public safety. That much is not disputed. The plaintiffs have not challenged the decision to require all officers to have the physical ability to perform the duties of a patrol officer. The contention made is that the PAT does not measure that ability. That is, that the PAT does not correlate with actual job performance.

The duties of a patrol officer are listed in the Department's job description, Defendant's Exhibit A-11. That describes a broad range of activity. The defendant has not shown how often the most physical activities are actually performed and what levels of strength and endurance the officer must exert.

The City relied on HPS to identify the essential duties of a police officer. HPS conducted a job task analysis for that purpose and as the first step of the test development. Ex. 1 ("Job Analysis"); Tr. Vol. VI (Eells) at 536:15-23; Tr. Vol. VII (Gebhardt) at 801:3- 804:6. That process involved collecting information about the tasks performed by incumbents in each rank, identifying essential tasks, and assessing the relative importance of each task and the frequency with which the tasks are performed. Physical abilities required to perform physical tasks were identified. Ex. 1, App. AA & AB. The physical abilities identified were muscular strength, muscular endurance, explosive strength, trunk strength, aerobic capacity, flexibility, equilibrium, and anaerobic power.

Using the results of the job analysis, HPS identified 12 physical skills tests that might be used as potential predictors of performance of essential physical tasks and the physical demands required to perform them. Ex. 2 at CSPD-PAT 00375-378.

HPS then conducted a "validation study" to assess the relationship between performance on the physical skills tests and job performance. For the study, HPS developed two "criterion measures" to use as the standards for job performance – (1) a work sample, and (2) a supervisor/peer rating form. Ex. 2 at CSPD-PAT 00379 -80 and App. A & App. B; Tr. Vol. VI at 534:12-537:11. The work sample was a timed exercise that consisted of a set activities related to pursing and restraining a suspect, such as getting out of a patrol car, running stairs, retrieving a gun, jumping over an obstacle, dragging a mannequin, and simulating an arrest. Ex. 2 at CSPD-PAT 00379 & App. A. The supervisor/peer rating form was a questionnaire for ranking an officer's performance in three categories: (1) physical job tasks; (2) physical abilities, and (3) overall job performance. Ex. 2 at CSPD-PAT 00380 & App. B. The form listed eleven physical job tasks and eight physical abilities to be evaluated, using a 7-point rating system.

Volunteers from the Department were recruited to participate in the validation study. One hundred seventeen officers (94 men and 23 women) were chosen. In 2011, the twelve physical skills tests that HPS had identified as potential predictors of performance were administered to the study participants. The participants also completed the work sample exercise, and supervisor/peer rating forms were distributed to their supervisors and peers.

HPS compiled the results from that physical skills testing and information obtained from the criterion measures, and analyzed the correlation between performance on the physical skills tests and the criterion measures. The push-up test, the sit-up test, the BEEP test and the Illinois

agility run were selected as the four components of the PAT. HPS determined the scoring system for each of those four components and established the overall cut-off score.

Reliance on an employment test that disparately impacts members of a protected class is "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." *Ass'n of Mexican Am. Educators v. California*, 231 F.3d 572, 584 (9th Cir. 2000) (quoting *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 431 (1975)). "'Validation' is the process of determining whether a selection device is sufficiently job related to comply with the requirements of Title VII." *Birmingham Fire Fighters Ass'n 117 v. Jefferson County*, 290 F.3d 1250, 1252 (11th Cir. 2002). A validation study may be used to show that an employment test is job related. *See* 29 C.F.R. § 1607.5A.

The study that HPS performed for the City was a criterion-related validity study. In *Ernst v. City of Chicago*, 837 F.3d 788, 796 (7th Cir. 2016), the United States Court of Appeals for the Seventh Circuit explained that a criterion-related study measures a study's validity by comparing the assessment tool (such as the skills tests scores) results with the criteria (such as the job performance rating or work sample scores). "If there is a strong correlation, the assessment tool is validated." *Id.*

To be valid, the study must accurately measure what it sets out to measure. When HPS developed the PAT and conducted the validation study, HPS was aware that the City intended to use the test not only as a selection tool but also for the assessment of incumbent officers. Pls.' Ex. 30; Tr. Vol. VI (Eels) at 541:13-20; Tr. Vol. VII (Gebhardt) at 792:13-18. HPS represented

that the physical abilities test it developed could be used for various purposes, including *as a component* of employment decisions such as selection, retention, and promotion. Pls.' Ex. 30; Tr. Vol. VII (Gephardt) at 828:9 - 829:34. The test was not designed to evaluate an officer's overall suitability for duty. The record evidence does not indicate that the PAT was developed for use as the *sole criterion* for termination of continued employment of incumbent officers with years of experience.

Importantly, HPS did not purport to determine the amount of physical fitness required to perform the various physical tasks identified as essential to the job. Tr. Vol. I (Weltman) at 16:2 - 17. Contrary to the City's argument, the record evidence does not show that the PAT tests an officer's ability to perform the minimum amount of physical activity necessary to effectively and safely perform the job.

Dr. Arthur Weltman, a professor of kinesiology and exercise physiology, testified on behalf of Plaintiffs regarding the PAT and the methodology underlying the HPS Job Analysis and Validation Report. He is qualified by education and experience to express opinions on those subjects. Dr. Weltman concluded that there was insufficient validity evidence to support use of the PAT. Tr. Vol. I (Weltman) at 85: 23 - 86: 3. His testimony is credible and persuasive.

Dr. Weltman opined that the push-up test, the sit-up test and the BEEP test do not measure the physical abilities that the PAT purports to assess with accuracy. HPS's job analysis ranked physical abilities required for the job and identified muscle strength as the highest. Tr. Vol. I (Weltman) at 40:3 -41: 10; Ex. 1 at CSPD-PAT 000058. According to Dr. Weltman, the one-minute sit-up and push-up tests measure muscle endurance rather than muscle strength. Tr. Vol. I (Weltman) at 41: 18 - 24; 51:12 - 53:1. Dr. Weltman said that those tests were not

appropriate for assessing a person's ability to perform tasks for which muscle strength is the most significant physical ability. *Id.*

The BEEP test was included in the PAT as an assessment of aerobic capacity, considered to be important for chasing a suspect. *See* Validation Report at CSPD-PAT 00445 (stating that BEEP test scores are representative of aerobic capacity). The BEEP test is a shuttle run that requires the participant to run 20 meters before a beep sounds, pivot and run another 20 meters in the other direction before another beep sounds, and then continue that process. As the test progresses, the beeps sound at shorter and shorter intervals, requiring the participant to run at a faster and faster pace. Tr. Vol. I (Weltman) at 27:1-8. Dr. Weltman testified that the BEEP test does not provide an accurate assessment of an individual's aerobic capacity. *Id.* at 28:7 - 34:5. He opined that the BEEP test is not a useful test for assessing a police officer's ability to run in pursuit. *Id*. at 35:5 - 22.

HPS's data from the study results showed that women officers' scores on the BEEP test were, on average, 82.15% of the men's scores. Validation Report at CSPD-PAT 00391. HPS also reported that the difference between men's and women's performance on the push-up test was significant, stating that women's score on the push-up test was 63.51 % of the men's score. *Id.* at CSPD-PAT 00392. Those skills tests favor men and do not accurately assess the physical abilities that they purport to measure.

Dr. Kraiger reviewed HPS's work and opined that HPS did not use an appropriate method for determining the cut-off scores for the PAT. Kraiger Ex. 2 at p. 3. He observed that using criterion measures to set passing scores is problematic. The evidence presented at trial bore that out. The criterion measures developed by HPS are of questionable value. The

Supervisor/Peer ratings are subjective. HPS did not attempt to validate the work sample. HPS did not analyze whether there was any correlation between the two criterion measures.

Most significantly, the evidence presented at trial revealed that the scoring system and cut-off score selected by HPS are meaningless.

HPS generated a point scoring system for each of the four component tests through a process that involved creating and analyzing sets of data described as "expectancy tables," "contingency tables," and "passing rate tables." Validation Report at CSPD PAT 00439-460; Tr. Vol. I (Weltman) at 68:17 - 76:12; Tr. Vol. VII (Gebhardt) at 812:2 - 20. As part of that process, HPS ranked study participants' performance on the physical skills test and then evaluated how certain levels of performance compared to satisfactory or unsatisfactory job performance. HPS defined satisfactory and unsatisfactory job performance according to a "composite job performance criterion measure," using results from the work sample and ratings from the supervisor/peer rating responses. In connection with that analysis, HPS decided that for the work sample results, a score of one standard deviation below the mean was classified as minimally acceptable. Validation Report at CSPD PAT 00448. HPS also decided that on the supervisor/peer rating form, any task or ability rating of 2.99 or below would be deemed unsatisfactory, although the 7-point rating scale on the form described a rating of "3"as "average" performance and "1" as "fair" performance. Tr. Vol. VII (Gebhardt) at 864:3 - 11; Validation Report at CSPD-PAT 00447.

Those judgments by HPS were arbitrary. Dr. Weltman explained that HPS's decision to use one standard deviation below the mean as the standard for minimal performance on the work sample meant that 16 percent automatically failed, without any indication of whether or not they

were able to perform the essential duties of a police officer. Tr. Vol. I (Weltman) at 74:22 - 76:12. Dr. Henderson acknowledged that HPS made an arbitrary judgment call, explaining that for the work sample there was no scientific way to determine satisfactory or unsatisfactory performance. Tr. Vol. V (Henderson) at 483:2 - 485:2.

With respect to the supervisor/peer questionnaires, the supervisors and peers who completed those forms were not informed that a below average rating would signify unacceptable performance. The form did not have a numerical rating for "unsatisfactory" performance. On cross-examination, Dr. Henderson suggested that it was reasonable for HPS to assume that the supervisor/peer ratings were inflated. Vol. V (Henderson) at 481:2 - 482:17. That testimony is an after-the-fact effort to justify a flawed validity study.[4]

To evaluate its scoring system, HPS applied the multiple hurdle approach and the compensatory approach to the validation sample. Validation Report at CSPD-PAT 00461 - 95. For the multiple hurdle approach, a cut-off score is established for each component skill test and a participant's failure to pass any one component constitutes failure of the entire test. For that approach, HPS set the passing scores at 28 sit-ups; 20 push-ups, and 20.47 on the Illinois Agility Run. *Id*. at CSPD-PAT 00445 - 447. The BEEP test has a two number score, signifying levels and intervals. HPS initially determined that 5,8 should be the passing score for the BEEP test. *Id*.

[4]When plaintiffs' counsel directed Dr. Henderson's attention to the contingency table for the push-up results, Dr. Henderson stated, " ... this is the part that I said was nonsensical." *Id.* at 494:2 - 16.

When HPS evaluated the sample test data according to the multiple hurdle approach, a high percentage of individuals with acceptable job performance failed under that approach. Validation Report at CSPD-PAT 00467.

HPS instead selected the compensatory scoring method. Validation Report at CSPD-PAT 00467 - 468. With a compensatory scoring method, a higher score on one or more component will compensate for a lower score on another component of the test. Tr. Vol. VII (Gebhardt) at 810:15 - 814:4.

The use of the compensatory scoring method to avoid failing high numbers of officers with acceptable job performance demonstrates that performance on the PAT's component tests does not correlate to job performance. The Validation Report's discussion of the scoring approaches acknowledges that an officer's failure to perform well on the PAT may not indicate inability to perform the job. Validation Report at CSPD-PAT 00467 - 468 (stating "high percentages of substandard performance [for the multiple hurdle scoring approach] may not be an accurate indicator of actual job performance."); Tr. Vol. I (Weltman) at 68:17 - 87:17.

For the compensatory scoring method, HPS determined that a passing score on the PAT required a composite score of 20, with at least one point on each of the tests four components. The City's own witness, Dr. Henderson, testified that the cut-off score established by HPS "has no relevance at all" and is "meaningless." Tr. Vol. V (Henderson) at 497:9 - 22.

When Dr. Kraiger reviewed HPS's scoring methodology, he detected a flaw in how the BEEP test was being scored. *See* Kraiger Ex. 2. HPS has acknowledged that error. That mistake is symptomatic of the flawed scoring system developed by HPS.[5]

[5]As a result of that change, an officer who previously had been judged to fail the test received a passing score. Tr. Vol. VII (Gebhardt) at 813:23 - 814:4.

In the Validation Report, the data regarding the "valid fail rate" percentages indicate that for any person who failed the PAT, the likelihood that their job performance was unacceptable was only 30.77%. Validation Report at CSPD-PAT 00469, Table 42; Tr. Vol. VII (Gebhardt) at 49:12-14. Dr. Weltman explained this means that for purpose of predicting performance, the PAT was inaccurate 7 out of 10 times. Tr. Vol. I (Weltman) at 87: 16.

Dr. Weltman found the correlations of the work sample to the PAT test scores to be low. *Id.* at 116:8-11. Dr. Weltman also found correlations between the Supervisor/Peer Ratings and the individual PAT tests to be low. *Id*. at 66: 7-12. Dr. Weltman concluded that there was insufficient validity evidence to support use of the PAT. Tr. Vol. I (Weltman) at 85: 23 - 86: 3.

Dr. Gebhardt said that the PAT has a high (.81) relationship with job performance. Tr. Vol. VII (Gebhardt) 809:1-24. That testimony is not persuasive. The criterion measures developed by HPS are not reliable assessments of job performance, and the cut-off score set by HPS is an arbitrary score.

The City emphasizes that HPS had expertise in designing physical abilities tests and validation studies, that the HPS study was extensive and its report contains voluminous analysis. These facts do not justify reliance on HPS's work. An HPS study was found invalid in other litigation. *See, e.g., Ernst v. City of Chicago*, 837 F.3d 788, 802 (7th Cir. 2016) (finding that validity study prepared by Gebhardt was faulty and not sufficient to show that City of Chicago's physical skills testing of paramedic applicants was job-related).

The fact that officers who failed the test were given multiple opportunities to pass it does not relieve the City of showing that the PAT is a valid test.

The fact that the overall passing rate was high does not show that the PAT is valid.

Ordinarily a court may not substitute its judgment on an employer's decision as what is an appropriate job requirement. A physical ability requirement may be reasonable for selection of new employees if it does not impose a barrier to that opportunity for any group protected by Title VII.[6]

To retroactively impose that requirement on women who have invested their lives as career police officers is fundamentally unfair. That is not to say that there can be no fitness requirement to maintain employment but to use physical tests that are not valid measures of the level of fitness that job duties actually require is a violation of Title VII when, as here, there is a disparate impact on women officers.

For the reasons stated the plaintiffs have prevailed on their claim that requiring them to pass the PAT to maintain their employment with the Colorado Springs Police Department violates Title VII.

Based on the foregoing, it is

DECLARED that the Colorado Springs Police Department's employment policy of using the physical abilities test designed by Human Performance Systems, Inc. as the exclusive standard for determining whether an incumbent officer is fit for regular duty violates Title VII of the Civil Rights Act of 1964, as amended.

Date: July 12, 2017

BY THE COURT:

S/Richard P. Matsch

_____

Richard P. Matsch, Senior Judge

[6]As indicated earlier, the PAT was a barrier for women recruits.