

411 F.Supp.2d 524
United States District Court,
W.D. Pennsylvania.

UNITED STATES of America, Plaintiff,
v.
CITY OF ERIE, PENNSYLVANIA, Defendant.

Civil Action No. 04-4 Erie.
|
Dec. 13, 2005.

Synopsis

Background: Federal government commenced suit against City of Erie, Pennsylvania, alleging that its use over several-year period of physical agility test (PAT) as device to screen police officer candidates violated Title VII in that it had disparate impact on female applicants and was neither job related for position of entry-level police officer nor consistent with business necessity. After city conceded that its use of PAT caused a disparate impact on female candidates, bench trial was held on remaining issues.

Holdings: Setting forth its findings of fact and conclusions of law, the District Court, [McLaughlin, J.](#), held that:

[1] to extent testimony of the three expert witnesses presented by federal government contradicted that of city's expert, district court credited the former over the latter;

[2] standard-setting exercise performed by nineteen incumbent police officers did not establish that PAT's 90-second passing score corresponded to minimum level necessary to successfully perform job of police officer;

[3] city failed to prove by preponderance of evidence the job-relatedness or validity of PAT as unitary test;

[4] city also had not established that its use of PAT with 90-second cutoff score was consistent with business necessity; and

[5] city had failed to rebut presumption that its use of PAT constituted unnecessary barrier to employment and was liable for violating Title VII, with specific relief to be determined in another phase of proceedings.

Judgment for plaintiff.

Procedural Posture(s): On Appeal.

West Headnotes (22)

[1] **Civil Rights** Educational requirements; ability tests

Labor and Employment Pre-Employment Issues

Term "validity" describes extent to which candidate's performance on employment test relates to his or her performance on the job; employment test is "valid" if candidate's test performance can be used to make a better prediction about how well he or she will perform on the job than might be possible without the test.

[1 Case that cites this headnote](#)

[2] **Civil Rights** Educational requirements; ability tests

Labor and Employment Pre-Employment Issues

A "content validity" study of employment testing must present data showing that content of the selection test represents important aspects of performance on the job for which test is being used; thus, "content validation" strategy requires evaluation of extent to which content of test is adequately matched to content of job to determine whether test measures what is important to or included within the job.

[3] **Civil Rights** Educational requirements; ability tests

Labor and Employment Pre-Employment Issues

Preemployment test can be judged to be "content valid" to extent that it represents contents of job, information about which is obtained from job analysis study; "content validation" strategy usually results in work-sample test or some type

of test that simulates the important aspects of job performance.

[4] **Civil Rights** Educational requirements; ability tests

Labor and Employment Pre-Employment Issues

In a “criterion-related validity” study, performance on employment test is compared with job effectiveness; criterion-related validity study should consist of empirical data showing that test is predictive of, or significantly correlated with, important elements of job performance.

[5] **Civil Rights** Educational requirements; ability tests

Labor and Employment Pre-Employment Issues

Employment test has “criterion-related validity” to extent that performance on test is statistically related to performance on the job; criterion-related validity is established when employer shows that scores on its test, even scores as basic as pass versus fail, related in a meaningful way with some measure of job performance, i.e., a criterion.

[6] **Civil Rights** Educational requirements; ability tests

Labor and Employment Pre-Employment Issues

The “construct validity” approach to employment testing is more theoretical than content or criterion-related validity because it is necessary to establish that a construct is required for job success and that the selection device measures that same construct; data from a “construct validation” study should show that test measures degree to which candidates have identifiable characteristics, i.e., constructs, that are important for successful job performance.

[7] **Civil Rights** Educational requirements; ability tests

Labor and Employment Pre-Employment Issues

The “construct validity” approach to employment testing requires showing that (1) test being validated measures a particular construct, and (2) construct is related to job performance; construct validity approach is most frequently used when it is well-established that particular test of a given construct, e.g., reading ability, has criterion-related validity, and researcher establishes that a different test, e.g., new reading test, also is valid by showing that the two tests measure the same construct, reading.

[8] **Evidence** Particular Subjects of Opinion Evidence

Evidence Opinion evidence

In suit involving physical agility test (PAT) which had disparate impact on female applicants for entry-level police officer positions in Erie, Pennsylvania, district court accepted as credible testimony presented by city's lay witnesses, most of whom were city police officers who testified about history of PAT, their own experiences taking test, and/or their own personal views about PAT's relevance vis-a-vis their own experiences as police officers; however, to extent those officers proffered their own personal opinions about usefulness, validity, or legality of PAT as employment screening device, their testimony was irrelevant as that was more properly subject for expert opinion. Civil Rights Act of 1964, § 703(k), 42 U.S.C.A. § 2000e-2(k).

[9] **Evidence** Conflicts between experts

In Title VII case against City of Erie, Pennsylvania involving physical agility test (PAT) which had disparate impact on female applicants for entry-level police officer positions, to extent testimony of the three expert witnesses presented by federal government contradicted that of city's expert, district court

credited the former over the latter. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

[10] Evidence  Scientific and technical works; learned treatises

Principles and Standards utilized in field of Industrial/Organizational Psychology and discussed by one of federal government's experts had relevance in Title VII suit against City of Erie, Pennsylvania for purposes of court's determination of whether physical agility test (PAT) having disparate impact on female applicants for entry-level police officer positions, as administered during six-year period, was job-related and consistent with business necessity. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

[11] Civil Rights  Physical requirements or limitations

Physical agility test (PAT) administered by City of Erie Pennsylvania, having disparate impact on female applicants for entry-level police officer positions, had to be validated as unitary test even though it consisted of obstacle course and push-up and sit-up components, since passing score consisted of but one requirement, completion of entire test within 90-second period. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

1 Case that cites this headnote

[12] Civil Rights  Physical requirements or limitations

While obstacle course portion of physical agility test (PAT) used by City of Erie, Pennsylvania to screen police officer candidates, if administered as separate test, would possess some unquantified degree of content validity, that finding was insufficient to establish the PAT's validity or job-relatedness in Title VII

disparate impact case. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

[13] Civil Rights  Physical requirements or limitations

Push-ups and sit-ups components of physical agility test (PAT) used by city of Erie, Pennsylvania for screening entry-level police officer candidates and found to have disparate impact on female candidates lacked content validity. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

[14] Civil Rights  Physical requirements or limitations

Push-ups and sit-ups components of physical agility test (PAT) used by city of Erie, Pennsylvania for screening entry-level police officer candidates and found to have disparate impact on female candidates lacked criterion-related validity. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

[15] Civil Rights  Physical requirements or limitations

Because City of Erie, Pennsylvania had not established that push-ups component or sit-ups component of physical agility test (PAT) for entry-level police officers measured construct of muscular endurance, City had not established that either component, as stand-alone test, had construct validity. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

[16] Civil Rights  Physical requirements or limitations

Standard-setting exercise performed by nineteen incumbent police officers in City of Erie, Pennsylvania did not establish that physical agility test's (PAT's) 90-second passing score for candidates for entry-level position corresponded to minimum level necessary to successfully

perform job of police officer; to the contrary, design of standard-setting exercise made it likely that passing standard was set too high. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

1 Case that cites this headnote

[17] Civil Rights  **Disparate impact**

Under the “disparate impact” theory of employment discrimination, Title VII prohibits employment practices which, although fair in form, are nevertheless discriminatory in operation. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

1 Case that cites this headnote

[18] Civil Rights  **Presumptions, Inferences, and Burden of Proof**

Burden upon employer of proving that their use of employment practice in question was job related and consistent with business necessity involves two-step inquiry in the Third Circuit, corresponding to each requirement. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

[19] Civil Rights  **Educational requirements; ability tests**

Standard for assessing the “business necessity” of a discriminatory passing standard or score on an entry-level police officer test is whether the passing standard reflects the minimum qualifications necessary to perform successfully the job in question. Civil Rights Act of 1964, § 701 et seq.,  42 U.S.C.A. § 2000e et seq.

1 Case that cites this headnote

[20] Civil Rights  **Physical requirements or limitations**

City of Erie, Pennsylvania failed to prove by preponderance of evidence the job-relatedness or validity of physical agility test (PAT) as it was

actually used for screening entry-level police officer candidates. Civil Rights Act of 1964, § 703(k)(1)(A)(i),  42 U.S.C.A. § 2000e-2(k)(1)(A)(i).

[21] Civil Rights  **Physical requirements or limitations**

City of Erie, Pennsylvania had not established that its use of physical agility test (PAT) with 90-second cutoff score which had disparate impact on female police candidates was consistent with business necessity. Civil Rights Act of 1964, § 703(k)(1)(A)(i),  42 U.S.C.A. § 2000e-2(k)(1)(A)(i).

[22] Civil Rights  **Physical requirements or limitations**

City of Erie, Pennsylvania failed to rebut presumption that its use of physical agility test (PAT) having disparate impact on female applicants for entry-level police officer positions constituted unnecessary barrier to employment and was liable for violating Title VII, with specific relief to be determined in another phase of proceedings. Civil Rights Act of 1964, § 703(k),  42 U.S.C.A. § 2000e-2(k).

Attorneys and Law Firms

***528** Benjamin Blustein, Christine M. Roth, Jay D. Adelstein, John M. Gadzichowski, Sharon A. Seeley, U.S. Department of Justice, Civil Rights Division (Employment), Washington, DC, Jessica Lieber Smolar, Mary Beth Buchanan, United States Attorney's Office, Pittsburgh, PA, for Plaintiff.

Gerald J. Villella, Dailey, Karle & Villella, Kenneth A. Zak, Paul F. Curry, City Solicitor's Office, Larry D. Meredith, Erie, PA, Katherine H. Fein, Pennsylvania Human Relations Commission, Pittsburgh, PA, for Defendant.

FINDINGS OF FACT AND CONCLUSIONS OF LAW

MCLAUGHLIN, District Judge.

This case was commenced on January 8, 2004 by the United States of America against the City of Erie based on the United States' allegation that the City's use of a physical agility test from 1996 to 2002 as a device to screen police officer candidates violated Title VII of the Civil Rights Act of 1964,  **42 U.S.C. § 2000e et seq.**, as amended. The United States alleges that the City's use of the test had a disparate impact on female applicants and was neither job related for the position of entry-level police officer nor consistent with business necessity. A non-jury trial was held from March 7 through March 10, 2005. Set forth below are this Court's findings of fact and conclusions of law:

I. FINDINGS OF FACT

A. Background Facts

1. For a number of years prior to 2004, the City of Erie administered a physical agility test ("PAT") as part of its process for hiring new police officers. The test was administered biannually on even years and each PAT remained in effect for a two-year period. (Tr. Vol. 1, pp. 159–63.)¹
2. In 1992, prior to the time period at issue in this litigation, the City of Erie ***529** utilized a PAT that incorporated pull-ups, push-ups, a vertical leap and a broad jump. (Tr. Vol. I, p. 161.) The pull-up component required male candidates to perform 4 pull-ups within 50 seconds and females to perform 1 pull-up within 50 seconds. (S 4.)
3. Due to a typographical error, the 1992 PAT was incorrectly administered, such that male candidates were erroneously required to do 6 pull-ups and female candidates were required to do 2. (S 5.) Because of this confusion, the City of Erie Civil Service Commission ² decided to allow candidates who failed the 1992 PAT to take the written portion of the exam and, if they passed, to then retake the pull-up component of the test as originally designed. (S 6.)
4. The Civil Service Commission's decision to allow candidates to retake the pull-up component of the 1992 test engendered some controversy. In the course of the
- controversy, the Civil Service Commission became aware of complaints that the pull-up component of the 1992 PAT was unfair. (S 7.)
5. As a result of the controversy arising out the 1992 PAT, the Civil Service Commission requested that the Erie Bureau of Police develop a new physical agility test. (S 8.)
6. Then Erie Police Chief Paul DeDionisio, Jr. assigned Stephen Kovacs the task of working with the Civil Service Commission to develop and document a new PAT that would be fair and more "state-of-the-art." At the time, Kovacs was Captain of the Bureau's Support Division, which oversaw training, research and planning for the Bureau. (S 9–11.)
7. Captain Kovacs had no prior experience developing physical agility tests; however, he and Chief DeDionisio reviewed tests used by other law enforcement agencies such as the Pennsylvania State Police and the Pittsburgh Police Department. Capt. Kovacs also contacted the Pennsylvania Municipal Police Officer Education and Training Commission and was informed that the Commission had not established a standard for physical agility testing and had chosen, instead, to leave the determination of those standards to the municipalities. (S 12–14.)
8. Captain Kovacs instructed Charles Bowers, then a Lieutenant in the Traffic/Patrol Division (now the City's present Chief of Police), to use his experience as a police officer to develop a test that would simulate physical tasks that city police officers regularly encounter. (S 15.) At that time, Lt. Bowers had been a police officer for more than 20 years and had extensive experience in patrol duties, including the pursuit and arrest of suspects or other persons who need to be restrained. (D 118.)³
9. Lt. Bowers had no education in the area of industrial/organizational psychology, exercise physiology, test development or test validation. (S 16.) Nevertheless, Bowers, like Kovacs and DeDionisio, understood that the City's reason for developing and administering the new physical agility test was to ensure that candidates possessed the physical ability necessary to do the job. (S 17.)
- *530** 10. Newly-hired City of Erie police officers are assigned to patrol duties, typically for a period of at least 5 years, and must work for promotions to non-patrol duties. (Tr. Vol. I, p. 202; D 90.)

11. In developing the new PAT, Lt. Bowers examined the physical tests utilized by law enforcement in various different cities, but he could find no national or uniform standard to work from. (Tr. Vol.I, p. 172.)

12. Based largely on his own extensive experience working the streets, Lt. Bowers attempted to construct a physical agility test which would simulate a foot pursuit containing common obstacles, followed by a demonstration that the individual taking the test had sufficient strength to apprehend and physically restrain a subject. (S 18; Tr. Vol. I, pp. 163–65.) In developing the PAT, Lt. Bowers did not rely on expert opinions or studies specific to the Erie police, nor did he conduct any formal study himself. (S 19.)

13. Lt. Bowers ultimately designed the PAT to consist of a 220-yard run, during which each applicant was required to negotiate four obstacles, followed by a push-ups component (to be completed immediately after the obstacle course/run) and a sit-ups component (to be completed immediately after the push-ups). The four obstacles included in the 220-yard obstacle course/run were, in order: a six-foot high wall, which applicants were required to climb over, a window opening three feet above the ground, which applicants were required to climb through, a platform two feet off the ground and eight-feet long, which applicants were required to crawl under, and a four-foot wall which applicants were required to climb over.⁴ (Tr. Ex. AA (Requests for Admission) and BB (Responses), Nos. 47, 48, 50, 53.)⁵

14. In March of 1994, based on his own experience as a patrol officer, Captain Kovacs recommended that the City use the PAT developed by Lt. Bowers. (S 21, 23.) In making this recommendation, Captain Kovacs did not rely on external academic or professional opinions or studies specific to the Erie Police Bureau, nor did he conduct any formal study himself. (S 24.)

15. Each of the obstacles in the obstacle course/run portion of the PAT developed by Lt. Bowers and recommended by Captain Kovacs was included because, based on the experience of Bowers and Kovacs, as well as informal discussions with other incumbent officers, each of the obstacles simulated something patrol officers would be required to do on the job. (S 26.) In developing and recommending the PAT, Lt. Bowers and Capt. Kovacs attempted to limit the components of the test to ones that simulated specific tasks that a police officer might encounter in a foot pursuit; for example, they did not include pull-ups

in the test because they did not believe pull-ups related to specific police officer tasks. (S 27.)

16. Captain Kovacs and Lt. Bowers included push-ups and sit-ups components in the version of the PAT recommended to the Civil Service Commission because they believed that push-ups and sit-ups, respectively, measure upper and lower body strength and, following completion of the *531 obstacle course/run, would indicate whether a candidate had sufficient strength or endurance to struggle with and apprehend a subject after a foot pursuit. (S 28.)

17. The PAT recommended by Bowers and Kovacs in 1994 initially did not specify the number of push-ups or sit-ups that would be included in the test or any duration for the push-ups or sit-ups components; instead Capt. Kovacs recommended that a “pre-test” of incumbent officers be conducted before the test was implemented in order to set the “exercise repetition standard” (i.e., number of push-ups and sit-ups included in the test). (S 29.)

18. Capt. Kovacs recommended that the number of push-ups and sit-ups required be set by “pre-testing” incumbent officers because he believed that: (1) the more physical ability (as measured by the push-ups and sit-ups) a candidate could perform, the better able to apprehend and subdue a subject that candidate would be; but (2) the City could not “require a higher physical ability level [than] that already demanded from active officers.” (S 30.)

19. Lt. Bowers chose 15 seconds as the time periods to be used for the push-ups and sit-ups components of the standard-setting exercise because he thought that a total of 30 seconds was about the duration of an average physical struggle. (S 31, Tr. Vol. 1, p. 172.)

20. In order to conduct the pre-testing or “standard-setting” exercise, Capt. Kovacs requested that Sgt. James Beskid, who was then an officer under his supervision, provide a list constituting a “random selection” of officers. (S 32.) Sgt. Beskid, who is not an expert in statistics or sampling, consulted a textbook and determined that a sample of at least 30 officers should be used for the standard-setting exercise. (S 33.) Sgt. Beskid produced a random list of 50 Erie police officers so that, in the event some officers were excused from participation in the standard-setting exercise, a sample of 30 would be available. (S 34.)

21. The Bureau of Police ordered the first 30 officers on the random list prepared by Sgt. Beskid to participate in the standard-setting exercise. (S 35.) However, the union representing the City's police officers took exception to the City ordering officers to participate without first negotiating with the union. (S 36.) Accordingly, the City issued a memo requesting volunteers from among the 50 officers on the list prepared by Sgt. Beskid, as well as other incumbent officers not on the list, to take the standard-setting exercise. (S 37.)

22. Ultimately, 19 incumbent officers volunteered to participate in the standard-setting exercise, which was conducted on May 19, 1994. (Tr. Vol.1, pp. 168–70, Ex. 11.) The use of 19 incumbent officers represented 14% of the total 134 Erie Police Officers assigned at that time to the Patrol/Traffic Division of the Erie Police Bureau. (D 89.) Of the incumbent volunteers, 3 were females and 8 were members of the SWAT team.⁶ (D 91, 95.) The average age of the incumbents was 35.8 years. (Ex. 11.)

23. In 1994, each of the 19 incumbent officers who participated in the standard-setting exercise was performing his/her job adequately. (S 39.)

24. The test was administered to the incumbents in such a fashion that each officer ran the 220-yard course, negotiating the four obstacles, immediately thereafter performing as many push-ups as possible *532 in a 15-second span and immediately thereafter performing as many sit-ups as possible in a 15-second span. The thirty-second allotment for combined push-ups and sit-ups was added to the officer's time on the obstacle course, to produce a final time. (Ex. 11.)

25. The average number of push-ups performed by the incumbents in 15 seconds was 17; the average number of sit-ups performed in 15 seconds was 9, and the average total time to complete the test was 87 seconds. (Ex. 11.)

26. Three female incumbents participated in the standard-setting exercise. Their respective scores are as follows. Candice McGahen: 15 push-ups/13 sit-ups and total time of 92 seconds; Tracy Stucke: 16 push-ups/8 sit-ups and total time of 86 seconds; Julie Kemling: 18 push-ups/11 sit-ups and total time of 80 seconds. (Ex. 11, D 93.)

27. The number of push-ups and sit-ups to be required in the PAT (as well as the cutoff time used as the passing standard) ultimately was the decision of the Civil Service Commission.

(S 40.) However, Lt. Bowers made recommendations as to the passing standards for the PAT.

28. Lt. Bowers recommended that the City use the average numbers of push-ups and sit-ups performed by the 19 incumbents in the standard-setting exercise, rather than their low scores, because he felt it would be fair. (S 41.) Lt. Bowers based this recommendation on the fact that the incumbent test-takers were, on average, older than most police officer applicants, and the incumbents (unlike new applicants) had not had time to prepare for the PAT. Lt. Bowers also believed that an officer's physical conditioning usually deteriorates once he or she is on the job, so he felt it was appropriate to hold new applicants to a standard higher than the lowest scores of the incumbent test-takers. (Tr. Vol.1, pp. 170–71.)

29. Lt. Bowers felt that using lower scores would not have met the needs of the Bureau. (Tr. Vol.1, p. 171.)

30. In setting the passing cut-off score for the PAT, Lt. Bowers advocated raising the time limit from 87 seconds (the average total time of the incumbents) to 90 seconds. This was not done for any particular reason, but simply to produce an even number. (Tr. Vol.1, p. 171.)

31. Capt. Kovacs consulted with Marcia Haller, an attorney who was then a member of the Civil Service Commission, regarding the number of push-ups and sit-ups that would be required (as well as the cutoff score or passing standard). Ms. Haller indicated to Capt. Kovacs that the City was not required to use the minimum performance by the incumbents and stated that they instead would use the averages. Ms. Haller believed the push-ups and sit-ups included in the PAT were a measure of "basic physical fitness." (S 42, 43.)

32. The representatives of the City involved in setting the PAT's passing standards decided to use the average numbers of push-ups and sit-ups completed by the 19 incumbents in the standard-setting exercise and to use the average time it took the 19 incumbents to complete the standard-setting exercise as the mandatory standard for new officer applicants. The City representatives chose those averages as a "medium" or "average" level of physical ability, believing that would be "fair" and "the best way to go." (S 44.)

33. Thus, as adopted and initially administered in 1994, the PAT required applicants to complete the 220-yard obstacle course, perform 17 push-ups and then perform 9 sit-ups, all within a 90 second period. Unlike the version administered to

the 19 incumbents (which consisted of three, separately time components), the PAT as adopted utilized a single passing ***533** standard of 90 seconds. (Tr. Ex. AA and BB, No. 59.)

34. Through the years the PAT underwent various permutations in an effort to maintain its relevance to the police officer job and to respond to complaints from the community regarding the number of women who failed. (D 98.)

35. In 1998, a 5-second grace period was added such that candidates who completed the test within 95 seconds were permitted one opportunity to take a re-test. (Tr. Vol. I, p. 174–76, Ex. 4.)

36. During the administration of the 1998 PAT, an 11-year old girl, described as “petite,” “wiry,” “very active” and a “gymnast,” observed the test and requested to take it unofficially at the end of the applicant testing. She passed all elements within the allotted 90 seconds. (Vol. I, pp. 75–77, D 117.)

37. Nevertheless, following administration of the 1998 PAT, local citizens representing the Erie County Human Relations Commission and women's advocacy organizations complained about the disparate passing rates and focused on the six-foot solid wall portion of the test as being excessively difficult for women and not adequately job-related to justify its use. (D 99.)

38. Accordingly, in 2000, pursuant to a recommendation by Chief DeDionisio and upon the Civil Service Commission's approval, the PAT was further modified such that applicants were given the option of climbing over *either* the six-foot wooden wall *or* a six-foot high chain link fence. Additionally, candidates had the option of using a 12-inch high wooden box for assistance. These changes were introduced in order to make the obstacles more appropriate and “practical” (i.e., like what an officer is likely to encounter in the City of Erie) and also to make the test more fair to women. (S 45, Tr. Vol. I, p. 175, 177, Ex. 9.)

39. Lt. Joseph Kress videotaped portions of the 2000 PAT administration and the tape was accepted into evidence as Defense Exhibit 1. (D 102.)

40. In 2002, the City introduced further changes to the PAT in an effort to increase the passing rate for females. (Tr. Vol. I, pp. 178–80.) For one, the PAT was administered only after applicants had first passed the written exam. Second, the push-up/sit-up components of the PAT were moved to the beginning of the test, such that they preceded the obstacle course. Third, the number of push-ups was decreased from 17 to 13, while the number of sit-ups was increased from 9 to 13.⁷ Finally, training sessions were scheduled and publicized for applicants. (Tr. Vol. I, pp. 178–80, D 105–106.)

41. Extremely wet weather on the date of the 2002 PAT resulted in a 5-second extension of the cut-off time to 95 seconds, with a further 5-second grace period being allowed on top of that such that candidates completing the PAT within 100 seconds could re-take the test once. (D 107.)

42. The changes in the 2002 PAT caused the rate of women passing to rise to 30% (7 of 23), so that those seven women, having passed the written test and the PAT, would be placed on the Civil Service list and ranked in order of their written scores, accounting for veterans' preference ***534** of 10 points as mandated by Pennsylvania law. (D 108.)

43. The City administered the PAT as part of each of the entry-level police officer selection processes between 1994 and 2002, the period under challenge here. Applicants had to pass the PAT in order to remain eligible for hire and continue on in the selection process. (Tr. Ex. AA (Requests for Admissions) and BB (Responses), Nos. 14, 21, and 22.) As noted, until 2002, the PAT was administered prior to the written exam.

44. The City of Erie Bureau of Police developed, and the Civil Service Commission approved and adopted the PAT, without consulting any expert(s) in the areas of physical abilities, job analysis, physical or other job requirements, employment testing or test validation. (S 46, 47.) No exercise physiologists or industrial/organizational psychologists were involved or consulted in the development of the PAT. (S 25.)

45. The following table represents the results of the City's use of the PAT between 1996 and 2002 in terms of male and female passing rates:

Year	Female Passing Rate	Male Passing Rate
1996	4.3%	53.7%

1998	14.3%	72.2%
2000	11.8%	77.3%
2002	30.4%	84.7%
1996–2002 Combined	12.9%	71.0%

(Tr. Ex. AA and BB, Nos. 27–31.)

46. As of the initiation of this lawsuit, the sworn workforce in the Erie Bureau of Police consisted of 193 men and nine women. At the time of trial, there were only eight female sworn officers (about 4% of the total sworn workforce) and only three female officers working in the patrol division. (See Complaint [Doc. # 1] at ¶ 14; Answer [Doc. # 3] at ¶ 1; Tr. Vol. 1, p. 197–98; *see also* Bowers Depo. Tr. (1/12/05) at pp. 21–22.)

47. In contrast, the police departments of other Pennsylvania jurisdictions, (i.e., Harrisburg, Pittsburgh and Philadelphia) have reported a percentage of female police officers earning at least \$25,000 per year in the range of 20–27%. (S 88.)

48. On August 5, 2004, the United States moved for summary judgment on, *inter alia*, the issue of whether the City's use of the PAT caused a disparate impact against female candidates. The City conceded the point and, on October 8, 2004, the Court found that the City's use of the PAT caused a disparate impact against female applicants for the entry-level police officer position between 1996 and 2002.⁸ (See 10/8/04 Hrg. Tr. [Doc. # 41] at pp. 16, 26.)

49. Between March 7 and 10, 2005, a bench trial was held on the remaining liability issues, *to wit*, whether the City's use of the PAT was job related for the position of entry-level police officers and, if *535 so, whether the passing standard used by the City was consistent with business necessity.

B. The Evidence Presented at Trial

50. The City presented the testimony of one expert witness, Paul Davis, Ph.D., and several lay witnesses, including current Chief of Police Charles Bowers, seven of the City's eight female police officers, and an instructor from the Mercyhurst College Municipal Police Training Academy.

51. The United States presented the testimony of three expert witnesses, David Jones, Ph.D., William McArdle, Ph.D., and Bernard Siskin, Ph.D.

Definitions/Background Relating to Expert Testimony

52. Industrial/organizational psychology ("I/O psychology") involves the application of principles of psychology and scientific research in the workplace, commonly in areas such as job analysis, job requirements, employment testing and selection, and performance measurement, among others. I/O psychology is recognized as a specific practice under Division 14 of the American Psychological Association. (S 55.)

53. In the field of I/O Psychology, there are published standards and principles that guide professionals in developing, using, and evaluating tests. Specifically, such standards and principles are stated in:

Standards for Educational and Psychological Testing (the "Standards") (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999); and *Principles for the Validation and Use of Personnel Selection Procedures* (the "Principles") (Society for Industrial and Organizational Psychology ("SIOP"), 2003)

(Tr. Ex. P at ¶ 16; Tr. Vol. 2, pp. 111–12; Tr. Vol. 2, pp. 71–72.)

54. The *Principles* focus specifically on the development, use and evaluation of personnel selection/employment tests. (Tr. Vol. 2, p. 71–72.) The *Principles* and *Standards* apply to the development and validation of physical tests. (Tr. Vol. 2, p. 110–112.)

[1] 55. The term "validity" describes the extent to which a candidate's performance on a test relates to his or her performance on the job. A test is "valid" if a candidate's test performance can be used to make a better prediction about how well he or she will perform on the job than might be possible without the test. (S 58.) Thus, professionals in

the field of employment testing and selection often use the terms “validity” and “job-relatedness” (or “valid” and “job-related”) synonymously. (S 57.)⁹

56. Validity is not an “all or nothing” concept; various tests may have different degrees of validity, and the same test may have different levels of validity for different jobs or when used in different manners. (S 59.)

57. Three types of validation strategies—content, criterion-related and construct—are described in the standards and guidelines followed by the employment testing profession. (S 60.)

[2] [3] 58. A content validity study must present data showing that the content of the selection test represents important aspects of performance on the job for *536 which the test is being used. Thus, a content validation strategy requires an evaluation of the extent to which the content of the test is adequately matched to the “content of the job” to determine whether the test measures what is important to or included within the job. A preemployment test can be judged to be content valid to the extent that it represents the contents of the job. Information about the content of the job is obtained from a job analysis study. A content validation strategy usually results in a work-sample test or some type of test that simulates the important aspects of job performance. (S 61.)

[4] [5] 59. In a criterion-related study, performance on the test is compared with job effectiveness. A criterion-related validity study should consist of empirical data showing that the test is predictive of, or significantly correlated with, important elements of job performance. A test has criterion-related validity to the extent that performance on the test is statistically related to performance on the job. Criterion-related validity is established when an employer shows that scores on its test (even scores as basic as “pass v. fail”) related in a meaningful way with some measure of job performance (i.e., a criterion). (S 62.)

[6] [7] 60. Finally, the construct validity approach is more theoretical than content or criterion-related validity because it is necessary to establish that a construct is required for job success and that the selection device measures that same construct. The data from a construct validation study should show that the test measures the degree to which candidates have identifiable characteristics (i.e., “constructs”) that are important for successful job performance. The user should

show empirically that the test validly relates the constructs to the performance of critical or important work behavior(s). This often requires a criterion-related study to show that the construct is related to job performance. Thus, a construct validity approach requires both: 1) a showing that the test being validated measures a particular construct; and 2) a showing that the construct is related to job performance. (S 63.) The construct validity approach is most frequently used when it is well-established that a particular test of a given construct (e.g., reading ability) has criterion-related validity, and the researcher establishes that a different test (e.g., a new reading test) also is valid by showing that the two tests measure the same construct (reading). (S 64.)

61. In his July 2004 report, Dr. Davis describes a “research design” validation method, which involves quantifying and comparing the metabolic (or energy) costs of job tasks on the one hand and of performance on the test on the other hand. (S 65.)

62. In his July 2004 report, Dr. Davis also refers to a “threat” or “perpetrator” analysis validation method, which involves building a model linking the demands of the test to the make-up of the City’s perpetrator population. (S 66.)

63. Whether a test is valid and whether a particular cutoff score or passing standard on the test corresponds to the minimum level of skills necessary to perform the job at issue successfully are two separate, though related, issues. (S 67.)

64. In the field of I/O psychology, professionally recognized methods by which cutoff scores are set appropriately include norm-referenced methods, content-related methods and criterion-related methods. (S 68.)

65. In his July 2004 report, Dr. Paul O. Davis, the City’s expert, also describes a “pacing” (or “concordant”), a “research design” and a “perpetrator” (or “threat analysis”) *537 method for setting or validating cutoff scores. (S 69.)

66. The “pacing” method of setting a cutoff score, as described by Dr. Davis, involves showing a group of subject matter experts (“SMEs”—i.e., individuals who are familiar with the job at issue—a videotape of incumbents or actors performing the test at various paces and then using a technique such as the Delphi method to determine the pace at which there is adequate agreement (or “concordance”) by the SME’s that the pace represents successful performance. (S 70.)

67. The “research design” method of setting a cutoff score, as described by Dr. Davis, involves selecting the level of performance on the test that requires a metabolic cost (or energy) cost that corresponds to the metabolic cost of performing job tasks. (S 71.)

68. The “perpetrator” method of setting a cutoff score, as described by Dr. Davis, involves determining the age and gender characteristics of the “perpetrator” population and selecting a level of performance on the test that corresponds to that required to apprehend some selected percentage of the perpetrators who flee/resist such that a foot pursuit and physical struggle is necessary to apprehend them. (S 72.)

Paul O. Davis, Ph.D.

69. Dr. Davis is a founder of Applied Research Associates, Inc., a research and consulting firm established in 1976. He has a doctoral degree from the University of Maryland, College of Health and Human Performance, Department of Kinesiological Sciences, where he placed major emphasis on the study of occupational fitness requirements and the quantification of work physiology. He has participated in over 60 legal proceedings as an expert witness, is certified by the American College of Sports Medicine, and has authored over 100 technical reports, manuals and articles dealing with his research on the relationship between human physical performance factors and health. (Ex. 17, p. 3.)

70. Dr. Davis was qualified at trial as an expert in the fields of exercise physiology, physical performance, physical requirements, and physical testing. Per the parties' stipulation, Dr. Davis is not qualified to provide opinion testimony in the fields of statistics, industrial/organizational psychology, testing other than physical testing, job requirements or job performance other than physical, and job analyses. (S 53.)

71. In connection with his services for the City, Dr. Davis reviewed a description of the PAT and viewed a videotape of portions of the 2000 PAT administration. On May 6, 2004 he visited the City of Erie, during which time he rode along with various officers on patrol and, in addition, interviewed the police chief and a number of officers. He was able to confirm through interviews that City of Erie officers routinely engage in strenuous physical activity in furtherance of their public safety mission. (Ex. 17, p. 5.)

72. Dr. Davis produced a July 14, 2004 “validation report,” which he characterizes as “validating the obvious.” It is the

result of his dissemination of a questionnaire to all 200+ Erie police officers, 114 of which were returned completed. The answers on the questionnaire indicate that nearly all of the responding officers had engaged in a struggle with a suspect or other person while on duty. (D 136.)

73. Dr. Davis testified at trial and his two reports (Ex. 16 and 17), as well as numerous excerpts from his deposition testimony, were admitted into evidence.

74. Boiled to its essence, Dr. Davis' opinion encompasses several propositions: (i) that the PAT which is the subject of *538 this litigation is a valid test of the physical demands of police work; (ii) that the PAT is not stringent enough to measure the amount agility, strength and endurance necessary to match the criminals who would be pursued in real-life scenarios; and (iii) that it is both unprincipled and dangerous to impose gender-normed standards for the physical screening of police officer candidates solely for the sake of achieving gender parity in the work force.

(i) Dr. Davis' Theory that the PAT is Valid

75. Dr. Davis opines that the challenged PAT used by the City of Erie is a valid test in that it measures physical abilities relevant to the successful performance of police work.

76. Specifically, Dr. Davis opines that the obstacle course portion of the PAT is content valid in that the obstacle course barriers are “exact replications” of the sorts of obstacles a police officer would commonly encounter in a foot chase. (Ex. 16, p. 3.) He therefore considers these general foot pursuit tasks as “self-evident and face valid.” (*Id.*; Ex. 17, p. 6.)

77. Dr. Davis further opines that the push-up and sit-up portions of the PAT have construct and/or criterion-related validity. (Ex. 17, p. 4; Tr. Vol. 2, pp. 78–79.) He believes that push-ups and sit-ups as performed on the PAT test for muscular endurance and are a “crude approximation of the energy costs of a struggle.” (Ex. 17, p. 4. *See also id.* at p. 6; Tr. Vol. 2, pp. 49–50, 84; Tr. Vol. 2, p. 78) (opining that the push-up and sit-up components have construct validity in that the metabolic demands of the muscle in performance of these evolutions involve the same muscle groups that are involved in altercations and struggles; thus they are an important construct of successful job performance).

78. He feels that the PAT appropriately utilized push-ups and sit-ups by placing them at the end of the test, following the obstacle course. (“Since this item is administered at the

conclusion of the test battery, a fatigue effect will clearly degrade maximum performance. For this reason, this is an appropriate time to interject the test item, since it reasonably reflects the struggle at the conclusion of a short sprint.” (Ex. 17, p. 6.)¹⁰

79. Dr. Davis’ theory presumes, as a fundamental proposition, that the PAT’s validity can be determined by examining the validity of each component part. While he agrees that one cannot validate the PAT as a whole simply by validating one component part alone, he believes one can validate the test as a whole by separately validating each of its component parts. (Tr. Vol.2, pp. 80–81.)

80. Dr. Davis did not perform or analyze any criterion-related validity study for the PAT as a unitary test because, he says, this would have been too costly. (Tr. Vol.2, p. 79.) In fact, he never performed any content, criterion, or construct validity study relative to the PAT as one whole test. (*Id.*)

81. Dr. Davis testified that there was no real need for a full-blown validation methodology such as was used in *Lanning v. SEPTA*, as the PAT is a much simpler, albeit gender-disparate test. Although there was no full-blown perpetrator analysis done here as there was in *Lanning*, Dr. Davis feels this is not a problem. Such a study would have been prohibitively costly in his view and, he claims, would not have *539 added any particular insight. (Tr. Vol.2, pp. 63–64.)

82. Dr. Davis concludes that the PAT is valid “by reason of its relationship to the essential functions of the job.” (Ex.16, p. 7.) The test, he believes, “contains relevant obstacles and an approximation of a minor scuffle that are as basic as one can get.” (*Id.*) (See Ex. 17, p. 10 (“The current and past physical ability tests are valid by virtue of their relationship to the occupational requirements of the job.”)) (See also Tr. Vol. 1, pp. 43–44.)

(ii) *Dr. Davis’ Opinion that the PAT Is Too Easy*

83. Dr. Davis opines that the PAT is too easy a test in that it does not adequately measure the amount of agility, strength and endurance necessary to match the criminals who would be pursued in the real-life scenario proposed by the PAT. (Ex. 17, pp. 7, 10; Tr. Vol. 2, pp. 45–46.)

84. Among other things, Dr. Davis criticizes the decision to modify the PAT such that candidates were given the option of scaling a chain-link fence in lieu of the 6-foot wall and the

option of using a crate or other scaling device. Dr. Davis views both of these modifications as an “erosion of the rigger of a selection instrument.” (Ex. 16, p. 3.) In Dr. Davis’ opinion, while technique and training both play a role in one’s ability to scale a wall, it is “far better to take individuals who present with no difficulty in performing this task rather than those who are struggling.” (Ex. 16, p. 3.) Dr. Davis believes the wall-scale is “powerful in its predictive value for the selection of applicants.” (*Id.*)

85. Thus, Dr. Davis believes that “more is better” (i.e. a higher cut-off is more preferable) when screening for physical attributes of police work, and he recommends the City restore its use of earlier versions of the PAT. (Ex. 16, p. 3.) (See also Ex. 17, p. 7 (“By manipulating the pass rates, the City may have unknowingly set in motion that trip down the proverbial slippery slope. Short answer, there is no precise point whereby failure is guaranteed. However, we do know that there is less risk (and no additional personnel costs) in increasing standards, particularly standards that reflect the profile of the perpetrators.”)) On the other hand, Dr. Davis concedes that this “more is better” approach is confined solely to the realm of physical job requirements, abilities and performance. (Tr. Vol.2, p. 77.)

86. Dr. Davis refers to the anecdote of the 11-year-old girl passing the PAT and states this “would certainly give any citizen pause as to the requisite rigor of this screening tool.” (Ex. 16, p. 3.) He believes that, if an 11 year-old girl can pass the test, then the test needs to be harder. (Tr. Vol.2, pp. 91–92.)

87. He estimates, based on prior experience as a gym teacher, that 60–70% of his students could have passed the PAT, at the same time acknowledging that kids are clearly unfit to be police officers. This indicates to him that the PAT is missing a vital ingredient: the strength component that is clearly demonstrated in terms of grappling with, engaging and subduing a perpetrator. (Tr. Vol.2, p. 67–68.)

88. Thus, Dr. Davis opines that the PAT is too easy because it doesn’t test some physical abilities or attributes (i.e., muscular strength) that are more important to police officer performance than the skills the PAT does measure. (Tr. Vol.2, p. 92.)

89. Dr. Davis also criticizes the PAT for its use of incumbents in setting the passing score. First, he believes that, because incumbents have tenure, they lack any incentive to

excel. Consequently, he believes the data derived from this approach is “contaminated with the ambiguities *540 of motivation.” (Ex. 16, p. 5.) Second, it is erroneous, Davis believes, to assume that all incumbents perform their jobs satisfactorily. (*Id.* at pp. 5–6.)

90. In addition, Dr. Davis believes that officers inevitably decline in fitness and agility from the moment of their hire. “Since advancing age will inevitably have a negative impact on performance,” he writes, “hiring those who are on the cusp virtually guarantees obsolescence in a fairly short period of time.” (Ex. 17, p. 7.)

91. Thus, while Dr. Davis criticizes the City's use of incumbent officers in structuring the 1994 standard-setting exercise, he supports the City's decision to utilize the incumbents' mean scores—rather than the lowest scores—as the relevant testing standard. (*See* Tr. Vol. 2, p. 54–55 (given age-associated decline in performance, the use of incumbents' mean score as a cut-off *does* represent the minimum qualifications necessary to do the job).) Setting standards based on the poorest performing incumbents, he believes, “is a strategy for disaster” and will never allow for improvement in the workforce. (Ex. 16, p. 6; Ex. 17, p. 7.)

92. Dr. Davis believes the better practice, rather than using incumbents, is to set physical standards based on the profile of criminals who will likely be the target of police action. (Ex. 16, p. 1 (“[I]t is the criminal element that defines the mission of the law enforcement officer.”); *id.* at p. 5 (“Better yet, the test should be an expression of success in the apprehension of criminals. The data on the criminal population supports the notion that they are younger than police officers.”); Ex. 17, p. 8 (“We are not arresting each other; we are supposed to arrest the bad guys. ‘What do they look like?’ is the more apt question.”); *id.* at p. 9 (“[F]itness standards need to be modeled on the basis of the threat.”); *see also* Tr. Vol. at p. 49.)

93. In this regard, Dr. Davis analyzed the booking information for over 2,300 Erie male arrestees,¹¹ and noted that 71% of those arrested in the year or more time span of the bookings are males under age 40, which he states is the population most likely to flee the scene or resist arrest. (D 136.)

94. He posits that the typical patrol officer is already giving the perpetrator the advantage of 10 years and 15 pounds of gear in a foot pursuit. “This fact alone,” he writes, “speaks to the need for above average levels of fitness.” (Ex. 17, p. 9.)

95. Dr. Davis opines that, while the PAT is valid, “[t]he test only modestly approaches what an officer may reasonably be expected to do and should be revisited with the eye towards providing a more reasonable representation of police work.” (Ex. 17, p. 10.)

96. For example, Dr. Davis does not believe the push-up/sit-up element of the PAT can be validly challenged as not job-related: “The expectation to perform a few push-ups and sit-ups at the conclusion of a brief sprint does not begin to approximate the physical demands of a real struggle in effecting an arrest.” (Ex. 16, p. 4; *see also* Ex. 17, p. 6 (“[T]he minimum number of push-ups and sit-ups hardly approaches thresholds that would have some predictive power.”).)

97. In this regard, Dr. Davis likens the PAT to the “big E” on the eye chart—“necessary but insufficient to establish visual acuity.” (D 136.) In other words, Dr. Davis believes the PAT as administered between 1996 and 2002 was so easy that, if a candidate could not pass it, this would necessarily indicate that the candidate was *541 not fit enough to perform the job of police officer and the City would be at risk in hiring that individual. (Tr. Vol. 2, p. 86.)

98. He opines that the PAT “is better than no test.” However, he believes that the City's use of the PAT placed it “at great risk of accepting individuals who cannot perform the rigors of the job since the test's action limits are well below those required on the job.” (Ex. 17, p. 7.)

99. Nevertheless, despite his view that the PAT is too easy, Dr. Davis believes it succeeded in at least screening for the minimum qualifications necessary to perform the job of police officer. (*See* Tr. Vol. 2, p. 89 (“I would hate to hire without the use of at least this by way of an instrument. So that to that extent this is a business necessity to have a test at least of this difficulty.”); Tr. Vol. 2, pp. 45–46 (“If I were to take fault.... I do not believe that this test is stringent enough. I do believe that it meets categorically the expectation of the minimum, minimum, minimum physical standard.”); Tr. Vol. 2, pp. 57–59 (opining that the PAT has “statistical noise” and “is not a very good test; it minimally, minimally meets the requirements to be a police officer,” but does not make him “warm and fuzzy to believe that a person who passes is always capable of handling every event.”)).

100. Dr. Davis specifically opined that the PAT's use of a 90-second time limit comports with the minimum qualifications necessary to do the job. (Tr. Vol. 2, p. 53.)

101. He feels that the videotape is “very compelling visually” as to the reasonableness of using the 90-second cut-off. He feels that the possibility of false negatives is very remote. (Tr. Vol.2, p. 60.)

102. In essence, Dr. Davis' opinion seems to be that the PAT could create false positives (i.e., individuals who are actually incapable of meeting the physical demands of police officer work might pass the PAT), but it probably does not create false negatives (i.e., any individual who failed the PAT *ipso facto* could not meet the physical demands of police work).

103. He believes that any reasonably motivated person of either gender should be able to pass the PAT. (Tr. Vol. 2, p. 68 (The test is “not out of reach of the physically inclined.”).)

104. He further opines that there is little likelihood for the substitution of an alternative approach that would have a lesser adverse impact. (Ex. 16, p. 7.) “The City has attempted to ‘dummy down’ this test and should return to its original version of the PAT. This slippery slope of tweaking the results to meet some social agenda benefits no one.” (*Id.*) (“Arguing in favor of reducing standards to achieve some agenda is contrary to the needs and interests of the citizens of Erie.” (Ex. 17, p. 9.))

(iii) *Dr. Davis' Opinion That the Test Should Not Be Gender Normed*

105. Dr. Davis believes that if different testing standards are employed for different people (e.g., gender-norming tests), then by definition the testing standards cannot be job-related, although they might possess differing degrees of job-relatedness. He believes that “you're undermining your position with regards to a defensible standard when in fact you have multiple standards for the same job.” (Tr. Vol.2, p. 83.)

106. While there may be wide-spread use of gender-norming standards in law enforcement with respect to calisthenics, etc., Dr. Davis believes that the more defensible approach is to identify what the “action limits” or requirements of the job are, independent of the officer's age and gender. He claims that the current trend is moving toward single standards. (Tr. Vol.2, pp. 96–97.)

***542** 107. Furthermore, Dr. Davis takes strong exception to the idea of coercing law enforcement entities to adopt double standards for men and women in the area of physical fitness and/or elevating characteristics such as “integrity” and

“sensitivity” over the critical fitness and agility necessary to perform police officer work. To do so, he believes, compromises law enforcement's standards and endangers the safety of police officers and the general public. Dr. Davis opines that “[t]o force the city to accept sub-par employees benefits no one. It is only the most misguided that somehow believes that placing unqualified people in harm's way to be [sic] a societal victory.” (Ex. 16, p. 6.)

108. He writes in his May 10, 2004 report that “[t]he wide range of physical tasks to be performed within law enforcement have no age or gender bias; that is, the job is not defined by the individual performing the job.” (Tr. Ex. 16, p. 1.) He further writes that, “Whenever there is a physical ability test that has any degree of sensitivity and specificity, the pass rates will be in favor of males,” and “[a]ttempting to adjust these rates to approximate an abstract definition of equality is fallacious.” (Ex. 16, p. 4.) He believes that “[t]he job is the job” and does not change as a consequence of who is performing it. (Ex. 17, p. 10; *see also* Tr. Vol. 2, p. 66.)

David Jones, Ph.D.

109. David Jones, Ph.D., is President of Growth Ventures, Inc., a human resources consulting firm with expertise in the design, validation and administration of employee assessment and selection systems. He holds a Ph.D. in the field of industrial/ organizational psychology and has practiced in that field for nearly 30 years. He has executed projects to design employee assessment and selection systems for both public and private sector organizations. (Ex. O, ¶ 1.)

110. Dr. Jones was qualified at trial as an expert in the field of industrial/ organizational psychology, including employment testing and selection, job analysis, physical and non-physical job requirements, and employment test validation. (Tr. Vol.2, p. 180.)

111. Dr. Jones testified at trial and his expert report was admitted into evidence as Ex. O. In addition, a rebuttal report co-authored by Dr. Jones and Leetta Hough, Ph.D., was admitted into evidence as Ex. P.

112. Dr. Jones rendered two central opinions: first, that the PAT does not show evidence of content, criterion-related, or construct validity sufficient to justify its use in screening entry-level police officer candidates; second, that the cut-off score utilized by the City is well above that which would be defined as the minimum standard of performance. (Ex. O, ¶¶ 8, 32–33; Tr. Vol. 2, p. 109.)

113. Dr. Jones testified that the *Principles and Standards* apply to the development and validation of physical tests. (Tr. Vol.2, pp. 110–12.) Thus, exercise physiologists typically draw upon APA Standards and SIOP Principles in determining whether a physical test is valid. (Tr. Vol.2, p. 112.)

114. Dr. Jones opines that Dr. Davis' reports bear no resemblance to validity studies described in APA Standards and SIOP Principles. In essence, Dr. Jones views Dr. Davis' reports as an expression of professional opinion unaccompanied by any type of validity study. (Tr. Vol.2, pp. 112–13.)

115. According to Dr. Jones, “content validity” is an appropriate strategy when the content of the test truly represents the relevant job functions. The term “face validity” as used by Dr. Davis is not a *543 term recognized by APA Standards or SIOP Principles. (Tr. Vol.2, pp. 127–28.)

116. Dr. Jones feels that it is not possible under professionally recognized standards to have criterion-related validity in the absence of empirical data demonstrating the relationship between test performance and job performance. As Dr. Jones explains, if one does not possess such data, one does not have a criterion-related study, and it would not be acceptable under professional standards to opine on the criterion-validity of a test based solely on professional judgment without supporting data. (Tr. Vol.2, p. 123.)

117. According to Dr. Jones, construct validity studies are a bigger and more theoretical undertaking; one must look at how a whole set of tests or a whole series of performance information fit together. (Tr. Vol.2, p. 124.)

118. In his preliminary report, Dr. Jones writes:

17. None of the information provided to me indicates that the Bureau's physical agility test has been held to content, criterion-related, or construct validation standards. While various changes have been made over successive administrations of the test, none of the work underlying these changes has been executed or documented in a manner consistent with professional practice.

18. Other than personal anecdote from the test's designers, there is no foundation on which to demonstrate the validity of the City's physical agility test, its individual

components, its scoring and qualifying standards, or its use in producing overall applicant eligibility lists. There is no information to show that individuals who perform well on the test perform similarly on the job.

19. In fact, no professionally sound job analysis has been conducted. There is no information to document whether each aspect of the job the City apparently attempts to simulate with the test are actually performed on the job. Nor is any information available regarding how frequently they might be performed, the specific physical capabilities that underlie their performance, or the degree to which new recruits can be trained to perform such activities once hired.

20. The City has made no attempt to assess the reliability with which the test is scored. For example, no test-retest analysis has been undertaken to determine the likelihood that candidates scoring within a given range of the cut-off score might succeed on a re-administration of the test, after a brief rest period. Given the test's degree of adverse impact, it would be reasonable for the City to determine the tool's test-retest reliability, to identify its standard error of measurement, and to offer retest opportunities to candidates scoring within a range of one to two standard errors of measurement.

21. No information has been provided to indicate that the City has sought out other, equally valid selection procedures that might produce less adverse impact on the employment opportunities of female applicants.

(Ex. O, ¶¶ 17–21.)

119. Dr. Jones thus opines that neither the process of initially designing the PAT nor the steps taken to modify it over the years were in conformity with professional standards. (Ex. O, ¶¶ 23–27, 28–31.)

120. Fundamentally, Dr. Jones opines that, because the PAT was administered as *544 a unitary test (i.e., the pass/fail decision was made based on the candidate's total performance of the test as a whole), it needs to be validated as a unitary test. (Tr. Vol.2, pp. 128–29.)

121. He explains that the PAT, as administered, is really a different test for every person who takes it because the amount of time required to complete each sequential segment in the test affects the amount of time the candidate has left to complete the remaining segments. (Tr. Vol.2, p. 130.)

122. Dr. Jones views this as significant because, in his words, “you have to know what you’re measuring in order to do a thorough study on whether a test has any validity or not.... If I run the obstacle course in 30 seconds, I have 60 seconds left [to complete the required push-ups and sit-ups.] If you run it in 60 seconds, you have only 30 [seconds]. At that point we’re now taking two different tests. And potentially measuring two different things from that point forward.” (Tr. Vol.2, pp. 130–31.)

123. One critical problem, in Dr. Jones’ opinion, is that the PAT standards were established in one fashion (i.e., as separately timed component parts) but administered in a different fashion (i.e., as a unitary test): “It’s putting together something that has the chance of being correct when it’s designed, but then putting it into place and making decisions in a completely different fashion.” These differences in administrative format can make a significant difference in the case of test measurement, according to Dr. Jones. (Tr. Vol.2, p. 131.)

124. Dr. Jones further opines that, even assuming it would be acceptable to demonstrate validity of the PAT by validating each component part separately, such validity has not been demonstrated here. (Tr. Vol.2, p. 133.)

125. Dr. Jones concedes that the obstacle course component of the PAT appears to have a degree of content validity; however, he finds no evidence of validity for the push-up and sit-up components. (Tr. Vol.2, pp. 133–37.)

126. As to the sit-ups component, Dr. Jones testified that he had never in his career seen evidence or studies showing that sit-ups relate to job performance for police officers. (Tr. Vol.2, p. 138.)

127. He believes it may be possible to construct a study which would evaluate the criterion-related validity *vel non* of sit-ups and push-ups; however, no such evidence has been gathered in this case. Moreover, two factors—the fact that an 11 year-old girl passed the PAT while incumbent SWAT team members ostensibly failed it—lower Dr. Jones’ expectation that the sit-ups and push-ups components of the PAT would have criterion-related validity under a proper analysis. (Tr. Vol.2, pp. 158–60.)

128. Dr. Jones also disagrees with Dr. Davis’ opinion that the PAT’s passing standards were set too low. In fact, Dr. Jones

opines that the standard was set far above that which would represent the minimum level of acceptable performance. (Ex. O, ¶¶ 32–33.)

129. Dr. Jones observes in his initial report:

32. The only systematic study of how current, effectively performing police officers might score on the test was undertaken in 1994. In that study, the performance of 19 current police volunteer officers was used to set a qualifying score on the test: the average score of the 19 officers on each component. While helpful in linking the qualifying score required of applicants to the performance of actual police officers, the “average officer” standard selected would have resulted in “disqualifying” approximately ***545** one-half of the current police participants.

33. Further, the “random sample” of officers on [sic] whom the passing requirement was established was not random. Originally, fifty Traffic Division officers were randomly sampled from a total list of 134 officers to participate in the process, with the objective of obtaining cooperation from at least 30. Only 19 officers volunteered and participated. This sample is not “random.” There is no basis for determining whether it reflected physical capabilities near the top, bottom or middle of the full officer force. It should be noted that some of the participants were actually members of the Bureaus [sic] SWAT team, a unit that has elevated physical ability requirements....

(Ex. O, ¶¶ 32–33) (internal citation omitted).

130. Dr. Jones further points out that, under the PAT standards as originally implemented in 1994, only four of the 19 incumbent test-takers—including only 2 of 8 SWAT team members—would have met all three criteria for a passing score (i.e., total time of 90 seconds or less on the PAT, 17 push-ups and 9 sit-ups). Under the standards as implemented in 2002, when the requisite number of push-ups was lowered from 17 to 13 and the requisite number of sit-ups was raised from 9 to 13, only 2 of 19 incumbents would have passed the PAT. (Tr. Vol.2, pp. 151–52.)

131. Dr. Jones considers this percentage passing rate “awfully low” considering that the City has represented that all 19 of the incumbent test-takers were adequately performing their jobs. (Tr. Vol.2, p. 152.)

132. Dr. Jones asserts this is all the more true in light of the fact that the incumbent test-takers were strictly volunteers who, in his opinion, tend to be more motivated and typically better performers. (Tr. Vol.2, pp. 152–53.)

133. Dr. Jones also observed that, with regard to the three female police officers who took and passed the PAT during its administration, their test scores reflect that they either passed by one second, failed by one second, or passed by virtue of a 5-second grace period. (Tr. Vol.2, pp. 154–55.)

134. Based on this evidence, Dr. Jones rejects Dr. Davis' "Big E" theory that the PAT is too easy. Dr. Jones concludes that this theory is belied by the fact that only three female officers (all admittedly competent) only barely passed the PAT, 30% of the males who took it failed, and a significant portion of the incumbents who set the test-taking standards would have received failing scores. (Tr. Vol.2, p. 157.)

135. Dr. Jones opines that, when conducting employment testing, there is a danger in setting the cut-off score too high on a single component, such as physical ability: namely, an employer risks eliminating those officers who might have an overall better mix of qualities and skills. Otherwise stated, Dr. Jones believes that the "job" is bigger than its physical requirements and, if the physical requirements are set too high, the employer misses its opportunity to add that blend of other relevant skills and abilities (such as cognitive abilities, job knowledge, personal character, etc.) into the mix. (Tr. Vol.2, pp. 155–56.)

136. With regard to the evidence of the 11 year-old girl who passed the PAT, Dr. Jones believes this factor does not necessarily mean that the PAT as such is too easy; in his opinion it more likely demonstrates that the PAT simply does not measure what the City intends it to measure. He agrees with Dr. Davis that the test probably should incorporate other components *546 that an 11 year-old likely would not pass. (Tr. Vol.2, p. 156.)

137. In sum, Dr. Jones believes that the PAT is an invalid test because some of its constituent parts (i.e., the sit-ups and push-ups) lack validity. (Tr. Vol.2, pp. 157–58.) In fact, Dr. Jones concludes that the City's methodology in designing, modifying, and scoring the PAT over time "fail[s] to incorporate even the most basic principles of content, criterion-related, or construct validation research" and that "[t]echnical documentation associated with the test design effort is nonexistent." (Ex. O, ¶ 35.)

138. Dr. Jones considers the procedures used by the City in scoring the PAT and in reaching its ultimate conclusions about applicants' qualifications to be "seriously flawed" in that they "bear no demonstrable relationship to the level of performance required for successful performance on the job." He further asserts that the scoring system "gives no consideration to the reliability of the test scores or to the degree that changes in the test's procedures over time have made the original standards-setting exercise obsolete." (Ex. O, ¶ 35.)

139. In Dr. Jones' opinion, an invalid test like the PAT should not be used because it is ultimately a waste of time and money and because it may produce unintended consequences, such as inadvertently screening out one group of job applicants over another. (Tr. Vol.2, pp. 141–42.)

140. Finally, Dr. Jones notes that, "[w]hile the City's attempts to reduce the level of adverse impact associated with the test are clear, they have been unrelated to any systematic analysis of the entry-level job, its physical requirements, or the existence of potentially equally valid, but less adverse alternative screening procedures." (Ex. O, ¶ 35.)

William McArdle, Ph.D.

141. William McArdle has a Ph.D. in work physiology and is Professor Emeritus in the Department of Family, Nutrition, and Exercise Science at Queens College in New York City, where he has taught since 1965. He has authored or co-authored more than 50 research articles dealing with environmental physiology, exercise training, and the assessment of diverse components of physical fitness as well as several text-books relating to exercise physiology. (Ex. V, p. 1.)

142. Dr. McArdle was qualified at trial as an expert in the field of exercise physiology. The parties have stipulated that Dr. McArdle is qualified to provide opinion testimony in the fields of exercise physiology, physical performance, physical requirements, and physical testing. (S 54.)

143. At trial, Dr. McArdle provided opinion testimony and his two reports were admitted into evidence as Exhibits V and W.

144. He was retained by the United States to assess the PAT's potential, because of physiological differences between the sexes, to cause an adverse impact on female candidates. (Ex. V, p. 1.)

145. Dr. McArdle opines that the agility run portion of the PAT represents a predominantly sprint-power or anaerobic form of exercise, as opposed to a more prolonged effort requiring energy from aerobic metabolism. Dr. McArdle notes that, while women generally perform less well than men in both anaerobic and aerobic physical tasks, the gender differences are “disproportionately magnified” in anaerobic tasks of relatively short duration. Thus, according to Dr. McArdle, “the overall test duration [of the PAT] with its high demand on anaerobic power capacity exacerbates the adverse impact on women.” (Ex. V, p. 2.)

***547** 146. Dr. McArdle notes that sit-ups and push-ups, to the extent they are used at all in the field of exercise physiology, are typically used to measure aspects of an individual's overall physical fitness. Moreover, tests of physical fitness use gender-specific standards in order to account for well-established physiological differences between men and women. (Ex. V, pp. 3–4.) Dr. McArdle therefore concludes that, to the extent the sit-up/ push-up portion of the PAT is used to measure candidates' physical fitness levels, it is inappropriate to use the same absolute performance requirements for both males and females. (Ex. V, p. 4.)

147. Due to physiological differences in upper-body strength, Dr. McArdle opines that the PAT's push-up component “would very likely disproportionately eliminate female candidates, even if the male and female candidates possessed the same general fitness level.” (Ex. V, p. 5.)

148. With regard to sit-ups, Dr. McArdle states that sit-ups do not reflect lower body muscular strength as the designers of the PAT believed. According to Dr. McArdle, “[t]he sit-up only assesses the endurance of the abdominal musculature. Considerable data exist to indicate gender differences, on average, [favoring males] in sit-up performances.” Thus, Dr. McArdle concludes, the PAT's requirement that male and female candidates perform the same number of sit-ups would “likely disadvantage female candidates.” (Ex. V, p. 5.)

149. Dr. McArdle concludes in his original report:

Based on my professional opinion as an exercise physiologist, the PAT's overall demands on sprint-power (anaerobic) physical capacity,

its reliance on upper-body muscle strength, and its failure to recognize well-established gender differences in measures of physical fitness magnify performance differences between men and women. The sit-up, wall climb, and push-up components have the greatest negative impact on a woman's chances for selection.

(Ex. V, p. 6.)

150. In his rebuttal report, Dr. McArdle disagrees with Dr. Davis' conclusion that the PAT is like the “Big E” on an eye chart—so easy that if any candidate who fails the test necessarily lacks the physical qualifications to be a police officer. (Ex. W, p. 3.) On the contrary, Dr. McArdle opined that the PAT's format—particularly where the sit-up/push-up component was placed at the end of the testing sequence—requires an “all-out physical effort” that represents a “predominantly sprint-power or anaerobic form of exercise, precisely the type used to produce exhaustion in exercise physiology laboratories.” (Ex. W, pp. 3–4.) (See also Tr. Vol. 3, pp. 89–90.) He points out that almost 90% of the women taking the PAT failed it; 25–30% of the males failed; Officer Raszkowski (a former state champion in the 2-mile run) barely passed the test, and 4 of 19 incumbents who participated in the standard-setting exercise would have failed it. (Ex. W, pp. 3–4; Tr. Vol. 3, pp. 90–91.) Dr. McArdle opines that “[t]he failure rate of applicants and incumbents on the PAT certainly indicates a difficulty level much greater than reading the top line of an eye chart.” (Ex. W, p. 4.)

151. Dr. McArdle does not believe that Dr. Davis' “Big E” theory is supported by the fact that an 11-year old female gymnast reportedly passed the PAT. Dr. McArdle points out that “[y]oung adolescent gymnasts, both male and female, typically possess an exceptionally high level of athletic prowess and physical fitness” and receive “specific training” which “includes vaulting and jumping maneuvers that would contribute greatly to success on the ***548** PAT.” (Ex. W, p. 4; see also Tr. Vol. 3, pp. 91–92.)

152. Dr. McArdle disagrees with Dr. Davis to the extent Dr. Davis suggests that failure to properly self-condition (because of lack of motivation and/or poor preparation) is the fundamental reason why women generally perform more poorly than men on the PAT. Dr. McArdle believes

that gender-specific differences between the sexes (including differences in body fat levels, *hemoglobin* concentration, muscular strength, and anaerobic and aerobic capacities) contribute dramatically to differences between the sexes in physiologic and performance capacities. (Ex. W, pp. 3, 4–5.) Thus, the fact that women do not perform as well as men on a test like the PAT does not indicate that they are less fit than or lack the similar motivation of their male counterparts. Dr. McArdle opines, for example, that it would be “ingenuous to consider that a woman who wins a marathon in a time 15% slower than a male counterpart is in some way ‘less fit’ or ‘lazier.’” (Ex. W, pp. 5–6.)

153. Dr. McArdle agrees that physiological differences between the genders do affect job performance in physical tests, but not to the extent reflected on the PAT scores because police work is highly skilled and involves complex direction of variables. (Tr. Vol.3, p. 95.)

154. Dr. McArdle disagrees with Dr. Davis' assumption that it is necessary to set the PAT cut-off score at a level considerably above the minimum requirement in order to account for inevitable age-related deficiencies in physical performance. Dr. McArdle asserts that there is research clearly indicating that loss of fitness is not an inevitable consequence of aging, particularly from young adulthood into middle age and depends more on lifestyle alterations that reduce physical activity than to biologic aging per se. (Ex. W, pp 3, 6; Tr. Vol. 3, p. 96.) Thus, Dr. McArdle maintains that “to select police officer candidates on the basis of ‘more is better’ so as to counter some anticipated aging effect fails to account for the behavioral effects on physical fitness” and “discriminates against the person with a lower (but acceptable) level of physical performance, yet whose lifestyle maintains this level throughout a career.” (Ex. W, p. 6.)

155. Dr. McArdle also criticizes Dr. Davis for ignoring the fact that physical performance of applicants may improve after hiring due to instruction and training on techniques for performing certain physical tasks. (Ex. W, pp. 3, 6.)

156. Dr. McArdle considers it “unjustified” from an exercise physiology point of view “to inflate a cut-score above the minimum requirement on the assumption that all individuals will lose considerable physical fitness as they progress in their careers and to ignore the fact that performance of job related tasks will improve with training.” (Ex. W, pp. 6–7.) Such a position, Dr. McArdle states, is “simply not supported by the professional literature.” (*Id.*)

157. Dr. McArdle disagrees with Dr. Davis' conclusion that the push-ups and sit-ups components of the PAT measure muscular endurance or that they are a useful measurement of an individual's ability to restrain a perpetrator. (Tr. Vol.3, pp. 79, 100.) He states that, when properly administered, push-up and sit-up tests “represent general physical fitness items that purport to measure the endurance of the muscles specifically activated in those movements.” (Ex. W, p. 7.) Dr. McArdle represents that typical physical fitness testing, as recognized by the American College of Sports Medicine, requires the individual to perform as many push-ups/ sit-ups as possible within a one-minute time frame or until exhaustion with no time constraint at all. He further states *549 that push-ups should not be linked to other events, and the format for women is typically modified, requiring them to push-up from the knees.¹² (Ex. W, p. 7, Tr. Vol.3, pp. 80–83.)

158. Dr. McArdle asserts that the push-ups and sit-ups as performed on the PAT—i.e., for approximately 15-second periods—provides no information on the endurance of the muscles involved in these activities or the ability to subdue and control a perpetrator. (Ex. W, p. 3; Tr. Vol. 3, pp. 79, 100, 125.) Dr. McArdle knows of no physical fitness test that infers muscular endurance utilizing as brief and variable a time period for sit-ups and push-ups as that utilized in the PAT. (Ex. W, pp. 7–8.) According to Dr. McArdle, it is not clear exactly what the PAT's sit-up/push-up components measure. The approximately 15-second time limits imposed by the PAT for the sit-up and push-up components is more a test of speed or power, depending on the amount of time each individual would have left himself to complete those components; however, it is not a test of endurance, nor does it represent any kind of standard measurement in the field of exercise physiology. A test of such short duration is not typically utilized to infer strength. (Tr. Vol.3, pp. 81–84.)

159. Dr. McArdle represents that he is aware of no professional literature relating the ability to perform 9 sit-ups and/or 17 push-ups in a 15-second time frame (or any other number in any other time period) to the ability to subdue or restrain a perpetrator. “This claim from Dr. Davis,” according to Dr. McArdle, “seems to emanate from a ‘Just believe me, I'm an expert’ point of view, without any objective data to support the claim.” (Ex. W, p. 3.) Thus, Dr. McArdle concludes that Dr. Davis' claim that one's ability to perform sit-ups and push-ups directly reflects ability to successfully overcome and/or restrain a combatant is not supported by empirical data. (Ex. W, p. 8.)

160. Indeed, Dr. McArdle considers it “naive” to believe that, just because the same muscle are activated in one physical activity versus another, performance capabilities become interchangeable. “Such logic implies that the person who lifts the most weights in the weight room will be the best shot-putter, high jumper, boxer, wrestler, or football player.” (Ex. W, p. 9.) “This is not the case,” Dr. McArdle explains, “because each of these sports, as with self-defense and restraint activities performed by police officers, requires use of muscles in highly specific and neural-controlled and learned movement patterns, which are not dictated simply by one's sit-up and push-up ability.” (*Id.*) This is the principle of “exercise specificity”—*to wit*, an individual must transpose his strength into the particular skilled activity at issue. (Tr. Vol.3, pp. 100–02.)

161. Dr. McArdle concludes that there is no evidence to support the proposition that the PAT’s 90-second cut-off score corresponds to a level of physical ability at or below the minimum physical fitness level required to adequately perform as an entry-level police officer for the City of Erie. (Ex. W, p. 10.)

Bernard Siskin, Ph.D.

162. Bernard Siskin holds a Ph.D. in statistics with a minor in econometrics from the Wharton School of the University of Pennsylvania and is former Chair of the Department of Statistics of Temple University. *550 (Ex. R, p. 1) Since receiving his Ph.D., he has specialized in the application of statistics to the analysis of employment practices and has been retained by numerous governmental and private organizations in that capacity. He has published numerous books and papers in his field, including articles on the role of statistics in the analysis of employment discrimination issues. He is currently the Director and Head of the Labor Practices Unit of LECG, LLC, in Philadelphia, Pennsylvania. (Ex. S, ¶ 1.)

163. Dr. Siskin was retained by the United States to provide an expert opinion as to whether statistical theory and evidence support Dr. Davis’ claims that the PAT is valid and whether the passing score was below the minimum level of physical skills required to do the job. (Ex. S, ¶ 3.)

164. Dr. Siskin is qualified to provide opinion testimony in the field of statistics and the application of statistics to employment practices and employment discrimination issues. (S 56.) He testified for the United States, and his expert report was admitted into evidence as Ex. S.

165. As set forth in his report of December 30, 2004, Dr. Siskin takes issue with Dr. Davis’s assertion that the City’s modifications of the PAT between 1996 and 2002, which have had the effect of “[r]educing the passing criteria [i.e. ‘dummifying down’ the test (*see* Ex. 16, p. 7)] does not ‘invalidate’ the test *per se*, but ... certainly reduces the sensitivity and specificity” of the PAT. (*See* Ex. 17, p. 7.) Dr. Siskin believes that, on the contrary, this statement by Dr. Davis concerning sensitivity and specificity supports a finding that the PAT is *invalid* rather than valid. (*See* Ex. S, ¶ 5.)

166. Dr. Siskin explains in his report that “sensitivity” refers to the percent of true negatives indicated by a test, while “specificity” refers to the percent of true positives indicated by a test. (Ex. S, p. 3.) Sensitivity and specificity are thus mathematically inverse to one another—e.g., lowering the cut-off score on a valid test will increase sensitivity but will decrease specificity, while raising the cut-off score will decrease sensitivity and increase specificity. Dr. Siskin therefore concludes that, in deciding where to set the cutoff score on a test that has some validity, the necessary trade-off between sensitivity and specificity should be considered. (Ex. S, p. 4.)

167. According to Dr. Siskin, if a test has significant “statistical noise” such that there is a low level of sensitivity and specificity (i.e., significant amounts of false positives and false negatives), then that test has little or no validity. (Ex. S, p. 4.) Thus, “[t]o the extent that Dr. Davis is of the opinion that the PAT has ‘a lot of noise,’ then both the sensitivity and the specificity fall and the test becomes similar to random selection and is deemed invalid.” (Ex. S, p. 5.)

168. Dr. Siskin testified that it is acceptable to validate a test as a whole by validating each component part separately if, and only if, two conditions are met: (1) the components of the test must be combined in such a way that they are structurally independent—i.e., an applicant’s performance on one component must not be affected by his or her performance on any other component(s); and (2) the components must not exhibit “multi-collinearity,” which, as Dr. Siskin explained, means that the components must be statistically independent, not correlated with each other. Stated simply, Dr. Siskin maintains that combining a valid test with a valid test does not necessarily produce a valid test unless these two conditions are met. (Tr. Vol.3, pp. 150–52.)

169. Dr. Siskin testified that the components of the PAT fail to satisfy either ***551** condition. (Tr. Vol. 3, p. 151–55, 169–70, 203–04; *see also* Ex. T-1, T-2, T-3.)

170. First, Dr. Siskin believes that the PAT components cannot be analyzed individually because they do not have structural independence: that is, they do not stand alone as separate, individual tests. Using Trial Exhibit T-1, Dr. Siskin demonstrated how a candidate's performance on the obstacle course run would determine the amount of time within which he or she would have to perform the subsequent components. Similarly, the length of time needed to complete the push-ups portion of the PAT would affect how much time the candidate would have left to complete the sit-ups portion. (Tr. Vol.3, pp. 151–52, Ex. T-1.)

171. Second, Dr. Siskin testified that, to the extent the City identified any statistical evidence of validity for any push-ups test, that evidence indicates that the components of the PAT are not statistically independent, but instead exhibit multi-collinearity. Specifically, Dr. Siskin testified that the only statistical evidence of validity for any push-ups test produced by the City consisted of the report of a study Dr. Davis conducted in 1988 of law enforcement officers in Anne Arundel County, Maryland. According to Dr. Siskin, that study shows only that the Anne Arundel push-ups test was a significant predictor of success at climbing a six-foot barrier. (Tr. Vol.3, p. 169.)

172. Dr. Siskin testified that this case presents an example of multi-collinearity: i.e., to the extent there is any evidence that a push-ups test would be valid as a stand-alone test (i.e. to predict a candidate's likely success in surmounting a wall/barrier), the PAT's push-ups component would not be statistically independent of the obstacle course/run, which already includes a six-foot barrier and thus directly tests the candidate's ability as to that task. As Dr. Siskin explained, individuals who successfully climbed the wall during the PAT's obstacle course/run but who did *not* successfully complete the PAT's push-ups component (and, therefore, are presumed *unable* to climb the wall) would be false negatives (i.e., incorrect predictions), and these false negatives would fall disproportionately to females. Thus, whatever validity the wall-climb obstacle might have standing alone would be reduced or eliminated by the addition of the push-ups test. (Tr. Vol. 3, pp. 152–56, 158–69, 169–170, 203–04; Ex. S, ¶¶ 26–33,35; Ex. T-2 through T-9.)

173. Dr. Siskin further opines that Dr. Davis is incorrect to the extent he suggests that increasing the passing score on the PAT will necessarily improve the *overall* performance of the City's police force. Dr. Siskin believes that whether an increase in the cut-off score will increase or decrease overall police force performance depends upon a number of factors, including the presence of other equally and more important job-relevant abilities. (Ex. S, ¶ 5.) Dr. Siskin explained that, assuming the PAT were a valid test, using a lower cut-off score would not necessarily equate to lowering the standards for police performance and using a higher cut-off score would not necessarily improve performance. Dr. Siskin asserts that there is a tradeoff: by increasing the requirements on the physical score, one necessarily downgrades the effectiveness on non-physical test scores, thereby making the non-physical skills less influential. Whether this results in a better or worse overall police force depends on where the cut-off is established and, thus, how this tradeoff is handled. (Tr. Vol.3, pp. 171–72.) Otherwise stated, increasing the cut-off score only makes sense, according to Dr. Siskin, if it eliminates those candidates who would not be able to do the job regardless of their other skills. (Tr. Vol.3, p. 172.)

***552** 174. Dr. Siskin testified that there is no statistical evidence (i.e., no criterion-related validity studies or supporting statistical evidence) allowing one to conclude that the PAT as a whole is a valid predictor of police performance or the physical aspects thereof. (Tr. Vol.3, pp. 147–48.)

175. Dr. Siskin further opines that there is no justification, from a statistical stand-point, for the pass/fail cut-off score utilized by the City on the PAT. (Ex. S, ¶ 5.) He concludes that the data presented does not support the use of the 90-second cut-off as being the minimum acceptable score. On the contrary, Dr. Siskin finds that the 90-second passing score is too high a standard based upon the methodology used by the City. (Tr. Vol.3, pp. 149–50.)

176. Dr. Siskin asserts that the City's standard-setting exercise (i.e. using the 19 incumbent officers) was unacceptable from a statistical standpoint. First, there were no measurements to establish how adequately those incumbents were performing their jobs. According to Dr. Siskin, this means that the City needed to perform an alternative type of norming study which required the City to first determine what percentage of the City's entire police force was not performing the job adequately from a physical standpoint (which the City did not do); then, the City would need a random sampling of incumbents to take the test and it could then adjust the cut-

off score accordingly, using a standard deviation. Dr. Siskin pointed out that, not only did the City fail to produce any estimation of the overall adequacy of its police force (from a physical task standpoint), but it also lacked a true random sampling of incumbent officers to participate in the standard-setting exercise. Dr. Siskin concludes that the sampling used here of 19 incumbents was tainted by the fact that it was based on volunteers, who are usually better performers. (Tr. Vol.3, pp. 177-80.)

177. In Dr. Siskin's opinion, the significant volunteer bias tainting the City's standard setting exercise is shown by the fact that 8 of the 19 volunteers were members of the City's SWAT team, who generally have to maintain a higher level of fitness. (Tr. Vol.3, p. 185-86.)

178. Dr. Siskin also notes that, when the performance of SWAT team members is discounted, the evidence shows that the non-SWAT incumbent volunteers performed nearly as well. Thus, the inclusion of the SWAT team members among the 19 incumbents did not dramatically affect the average scores achieved in the standard-setting exercise. Dr. Siskin views this fact as further evidence of the volunteer bias which infected the standard setting exercise. He theorizes that the non-SWAT incumbents performed nearly as high as the SWAT team members because the non-SWAT incumbents (being volunteers) were generally high performers by nature. (Tr. Vol.3, pp. 185-86.)

179. Dr. Siskin opines that, from a statistical viewpoint, the percentage of incumbents that would have failed the PAT is too high to suggest validity. (Tr. Vol.3, pp. 189-90, Ex. T-11.)

180. He further opines that the use of volunteers in the standard setting exercise makes the PAT invalid, absent adjustments that would take the effects of volunteerism into account. (Tr. Vol.3, pp. 194-95.)

181. Dr. Siskin explained that using a cut-off score of 1 standard deviation below the mean over time does not decrease the performance of the police force in terms of physical tasks, but rather increases it. (Tr. Vol.3, p. 182.)

182. Finally, Dr. Siskin opines that the City's utilization of females is well below *553 that of the police forces in Philadelphia, Pittsburgh, and Harrisburg. (Ex. S, ¶ 5.)

C. The Court's Credibility Findings

[8] 183. As a general proposition, the Court accepts as credible the testimony presented by the City's lay witnesses, most of whom are City police officers who testified about the history of the PAT, their own experiences taking the test, and/or their own personal views about the PAT's relevance vis-a-vis their own experiences as police officers. However, to the extent these officers proffered their own personal opinions about the usefulness, validity, or legality of the PAT as an employment screening device, the Court finds the testimony irrelevant, as that is more properly a subject for expert opinion and, ultimately, a determination to be made by this Court.

[9] 184. As a further general proposition the Court credits the testimony of the expert witnesses presented by the United States. To the extent their testimony contradicted that of Dr. Davis, the City's expert, the Court credits the United States' expert witnesses over Dr. Davis.

185. The Court finds that, while Dr. Davis is certainly well qualified to provide opinion testimony in the field of exercise physiology, in this particular case his opinions were not well supported by empirical evidence or persuasive analysis. Accordingly, the Court declines to accept his opinions concerning the validity and usefulness of the PAT, as set forth in more detail below.

1. Principles of I/O Psychology Have Relevance in this Case

[10] 186. The Court finds that the Principles and Standards utilized in the field of Industrial/Organizational Psychology and discussed by Dr. Jones have relevance in this case for purposes of our determination whether the PAT, as administered between 1996 and 2002, was job-related and consistent with business necessity.

187. Dr. Davis acknowledged that the American College of Sports Medicine (ACSM), an authoritative body in the field of exercise physiology, is presently initiating an action to produce its own set of principles and guidelines, but none have been established as yet. (Tr. Vol.2, p. 73.) Thus, Dr. Davis' own profession has no published standards for the development, use or evaluation of employment tests.

188. Because no standards have yet been published by the ACSM, the exercise physiology profession commonly borrows the standards utilized by the industrial/organizational psychology profession. (Tr. Vol.2, pp. 73, 112.)

189. Indeed, Dr. Davis claimed that his expert reports, setting forth his opinions and conclusions in this case, complied with the *Principles and Standards*. (Tr. Vol.2, p. 73.)

190. Despite Dr. Davis' representation that professionals in his field do not defer to the Society of Industrial/Organizational Psychologists or their *Principles* or *Standards* in conducting employment testing (see Tr. Vol. 2, p. 94), the Court is satisfied that the *Principles* and *Standards* have relevance in this case, particularly in the absence of other controlling testing standards in the field of exercise physiology.

2. The PAT Must Be Validated as a Unitary Test

[11] 191. The Court finds that the PAT, as administered by the City during the challenged time period, was a unitary test, since the passing score consisted of but one requirement—completion of the entire test within a 90-second period (95 seconds in 1998).

192. In previous proceedings, the City has conceded that the PAT as a whole *554 resulted in a disparate impact against females. The City did not argue that any of the PAT's components was a separate employment practice with respect to which the United States was required to prove disparate impact. (See Def.'s Br. in Resp. to Pl.'s Mot. for Summ. Judg. [Doc. # 36] at p. 3 (Dr. Siskin's "assertion of statistically significant gender disparity in the City of Erie's PAT is not in dispute.").) Since the "employment practice" in question which the City conceded, and this Court found, had a disparate impact was the use of the PAT as a single test, it follows that the City must prove that the PAT, as one test, was job-related and consistent with business necessity.

193. The component parts of the PAT cannot be validated independently because they are not structurally independent. That is to say, a candidate's performance on the obstacle course run clearly affects the candidate's performance on the subsequent components.¹³ Similarly, the length of time necessary to complete the push-ups determines how much time the candidate would have to complete the sit-ups portion. As an example, a candidate who takes 50 seconds to complete the obstacle course/run then has 40 seconds to complete the required numbers of push-ups and sit-ups, while a candidate who takes 70 seconds to finish the obstacle course/run has only 20 seconds for the push-ups and sit-ups. Similarly, how much of the 20 (or 40) seconds the candidate would have left to complete the required number of sit-ups depends upon how long the candidate took to complete the push-ups.

194. Thus, if each component of the PAT were to be considered separately, the "passing standard" would be different for every candidate who took the PAT, and it would therefore be unclear exactly what the City was attempting to justify as job-related and consistent with business necessity.

195. The Court therefore credits the opinions of Dr. Jones and Dr. Siskin who stated that, because of the way the PAT was administered to police officer candidates and the way it was used to make hiring decisions (i.e., as a single, unitary test), it must be validated as a unitary test. The Court declines to credit Dr. Davis' opinion that the PAT can be validated by validating each component part individually as if each were a stand-alone test.

3. The City Has Not Proved That the PAT as One Test is Valid.

196. The Court finds that the City has not proved that the PAT is valid as a unitary test in that it conducted no validation study on the PAT as a single test and Dr. Davis' unsupported opinions do not establish the requisite validity.

197. In setting the 90-second passing standard, the City did not perform or consider any professional analysis of the job tasks and duties of the City's police officer position or of the knowledge, skills or abilities required for successful performance in that position. (Ex. AA and BB, Request No. 76 and Response thereto.)

198. In determining the passing standard it would use on the PAT, the City did not perform or consider any validity study or analysis of the relationship between performance on the PAT and job performance. (Ex. AA and BB, Request No. 76 and Response thereto.)

199. Specifically, in setting the PAT passing standard, the City did not determine whether performing the PAT in 90 seconds corresponds to the minimum level *555 of ability that is required for successful performance as a police officer. (Tr. Vol.1, pp. 189–90.)

200. Rather, the City has stipulated that the representatives of the City involved in deciding to use 90 seconds as the passing standard for the PAT did so because they believed that the City would be requiring a "medium" or "average" level of physical ability, and that seemed "fair" and "the best way to go." (S 44.)

201. Dr. Davis performed virtually no research or analyses supporting his opinions about the PAT's validity, with the exception of a survey relating to only one of the three components of the PAT (the obstacle course/run).

202. Dr. Davis did not do a formal analysis of the PAT passing standard using the content, construct or criterion-related method. (S 82.)

203. Dr. Davis did not conduct a formal "research design" study to validate the PAT, nor did he conduct a "perpetrator" or "threat" analysis to validate the PAT. (S 75, 76.) Thus, while Dr. Davis suggested additional methods for establishing evidence of the PAT's validity, he never followed through on conducting any analysis under these methods. (See Ex. O, ¶¶ 45–50.)

204. In sum, Dr. Davis never conducted any validation study on the PAT as a unitary test.

205. Instead, relying on his education and experience, and utilizing what he termed a "common sense model" (see Tr. Vol. 2, pp. 36–37), Dr. Davis concluded that the test as a whole was manifestly content valid as to the obstacle course portion of the PAT and criterion and/or construct-valid as to the total time limit and muscular endurance measured by the push-ups and sit-ups. (S 82; *see also* Tr. Vol. 2, pp. 58. ("I trust my life experiences and my eyes probably far more than my academic training.").)

206. To the extent Dr. Davis claims that his opinion is based on his "experience," however, he has not identified any specific experience that could form a sufficient basis for that opinion. The City has stipulated that Dr. Davis has never validated, used or recommended a push-ups or sit-ups test like that incorporated in the PAT (i.e., a test of 15–seconds or similarly short duration), nor is he aware of any other organization that has ever used a test like the PAT. (See generally S 73–82.)

207. Dr. Davis' "Big E" theory assumes that the PAT validly measures an ability—including, apparently, muscular endurance and/or "fatigue resistance"—that is required for the job of police officer. Thus, Dr. Davis assumes one of the facts that Title VII requires the City to prove, *to wit*, that the PAT is valid. His opinion in this regard is not supported by any analysis or research and fails to establish that the PAT as a single test is valid. (See Ex. P, ¶ 152; *see also* Davis Depo. (11/17/04–11/18/04) at pp. 199–201) (stating that the "best evidence" that applicants who cannot do 17 push-ups in

15 seconds and 9 sit-ups in 15 seconds cannot successfully perform the police officer job is his "opinion that a person who couldn't minimally do this has a fairly poor prognosis of being a successful police officer."); *id.* at 197–198 (defining the terms "sensitivity" and "specificity").

208. Indeed, at times, Dr. Davis' testimony about the PAT was somewhat self-contradictory. Despite professing that the PAT has validity in screening out unqualified candidates, Dr. Davis acknowledged that it may not have strong predictive value. (See Tr. Vol. 2, pp. 88–89 (citing Dr. Davis' deposition testimony that, because the PAT is so easy, it really lacks strong predictive power).) In fact, in his *556 deposition, Dr. Davis opined that, "If the test is so incredibly insensitive to allow an 11-year old girl to pass it, it hardly rises to the definition of a valid test." (Tr. Vol.2, pp. 91–92.)

209. During his rebuttal testimony, after this Court had received testimony from the United States' experts explaining why the PAT must be validated as a single test, Dr. Davis did testify that the "entire" PAT measures "fatigue" or "resistance to fatigue." Thus, according to Dr. Davis, the PAT as a whole measures an individual's ability to perform a "standardized" or "fixed" amount of work within 90 seconds, the theory apparently being that those who cannot do the amount of "work" necessary to complete the PAT within 90 seconds lack the physical ability to perform the job. (See generally Tr. Vol. 4, pp. 7–9.)

210. This theory is, in essence, what Dr. Davis calls a "research design" approach to validity. The parties have stipulated, and this Court finds, that a research design validation involves quantifying and comparing the metabolic (or energy) costs of job tasks on the one hand and of test performance on the other. (S 65.)

211. However, as previously noted, Dr. Davis did not conduct a research design study on the PAT. Specifically, Dr. Davis never quantified how much work a police officer has to be able to do in 90 seconds to perform the job or any job tasks successfully. Nor did Dr. Davis establish how 17 push-ups and 9 sit-ups (or the obstacle course run) related to that amount of work. He never quantified the amount of work used in what he termed the "metric" of a push-up or a sit-up or the obstacle course run. (see S 73–74 (Dr. Davis did not quantify the metabolic costs of any City of Erie police officer job tasks, nor did he quantify the metabolic costs of the PAT or any of its components).)

212. Dr. Davis' fatigue theory does not address or answer why "push-ups" and "sit-ups"—as opposed to some other exercise—are the appropriate "metric" to measure energy costs of the job and why it is appropriate to demand 17 push-ups and 9 sit-ups as opposed to some other number. In short, Dr. Davis provided no evidence to support a finding that the ability to do the amount of work involved in the PAT in 90 seconds corresponds to the minimum level necessary to perform the police officer job successfully.

213. Although this Court fully acknowledges Dr. Davis' impressive credentials and extensive experience in the field of exercise physiology, in this particular case, his opinion cannot be credited and is insufficient to establish that the PAT is a valid test.

214. No one else has ever tried to validate the Erie PAT as a unitary test, nor has any other entity ever used it. (Tr. Vol.2, p. 79.)

215. The parties' frequent reference to the fact that an 11-year old girl passed the PAT also does not indicate that the PAT, as a unitary test, is valid. The Court accepts as credible the testimony of Dr. McArdle that it is unsurprising that an 11 year-old girl, particularly a petite girl trained in gymnastics, would be able to pass the PAT, given the fact that gymnasts are commonly trained in jumping and vaulting exercises, generally have lean body mass, and typically possess a significant amount of strength relative to their size. (Tr. Vol. 3, pp. 91–92; Ex. W, p. 4.)

216. In fact, the Court finds it noteworthy that Lt. Les Fetterman, a SWAT team member whom the City stipulates performs his job adequately (S 39), and whom the City contends is "one of [its] largest and strongest officers," (see D 120), failed to perform 17 push-ups in 15 seconds (he performed only 13) during the 1994 standard ***557** setting exercise, despite admittedly putting forth his best effort. (Ex. 11, Tr. Vol.1, pp. 139–40.)

217. Accordingly, the fact that an 11-year old gymnast reportedly passed the PAT does not prove that the PAT is an easy test for an adult, even an adult who possesses the physical abilities necessary to successfully perform the job of police officer. Rather, the Court credits Dr. Jones' opinion that this is more likely an indication that the PAT simply does not test all areas relevant to the physical work of police officers. This, in turn further supports the Court's finding that the PAT, as a unitary test, lacks validity.

218. The Court has viewed and considered the videotape (Def.'s Ex. 1) taken during portions of the 2000 administration of the PAT. However, the Court finds that the videotape possesses little evidentiary value for purposes of the ultimate issues in this case.

219. The videotape did permit the Court to visualize the obstacle course and push-up/ sit-up components. Furthermore, the tape did anecdotally demonstrate that there were officer candidates present at the 2000 administration of the PAT who appeared manifestly unfit to meet the rigors of the test. Nevertheless, the videotape is not a complete record of the 2000 test because only certain individuals were selected to be videotaped. Perforce, it cannot and does not prove that *no* candidates who took and failed the 2000 PAT were otherwise minimally qualified to perform the job of police officer for the City of Erie.

220. More importantly, we have already found that the *Principles* and *Standards* set forth above are relevant to this inquiry, and the videotape produced by the City provides no information to the Court as to whether the PAT is valid or consistent with business necessity under the *Principles* and *Standards*.

221. In summary, because the City has not established that the PAT as one test is valid, the City has failed to prove that the PAT is job related for the position in question, and the United States is entitled to judgment in its favor on the issue of liability.

4. *The City has not shown that each of the individual components of the PAT would be valid if administered as a separate test.*

222. Alternatively, even if the City's piecemeal approach to validation were appropriate—and we have specifically found it is not—the City has not shown that each of the components of the PAT would be valid if administered as a separate test.

223. The parties agree, as discussed above, that there are three types of validation strategies—content, criterion-related and construct—described in the *Standards* and *Guidelines* followed by the employment testing profession. (S 60.)

224. In addition, in his report, Dr. Davis describes two variations on these three accepted validation approaches. Specifically, Dr. Davis described a "research design" method, which, as noted above, involves quantifying and comparing

the energy costs of job tasks on the one hand and of performance on the test on the other, and a “threat” or “perpetrator” analysis, which involves linking the demands of the test to the make-up of the City's perpetrator population. However, the City has stipulated that Dr. Davis did not conduct either a “research design” study or a “perpetrator” analysis to validate the PAT. (S 65–66, 71–72, 75–76.)

225. Instead, Dr. Davis asserts that the obstacle course run component of the PAT has content validity, and the push-up/sit-up components each of criterion-related and/or construct validity. (Tr. Vol. 2, at 78–79; Ex. 16 at p. 3; D 135.)

558 a. *The obstacle course component of the PAT

226. The parties have stipulated that a content validity study must present data showing that the content of the test represents important aspects of performance on the job for which the test is being used. Thus, a content validation strategy requires an evaluation of the extent to which the content of the test is adequately matched to the “content of the job” to determine whether the test measures what is important to or included within the job. Information about the content of the job is obtained from a job analysis study. A test can be judged to be content valid to the extent that it represents the content of the job. A content validation strategy usually results in a work-sample test or some type of test that simulates the important aspects of job performance. (S 61.)

227. Dr. Davis asserts, based on his “job task analysis” survey, that the obstacle course portion of the PAT is a content valid simulation of a foot pursuit. (Ex. 17, p. 6.)

228. Although the United States' experts have maintained that Dr. Davis' “job task analysis” is flawed and does not constitute a professionally appropriate job analysis, the United States agrees that the obstacle course portion of the PAT, if administered as a separate test, likely would have some level of content validity, although the United States maintains this level is unquantified and likely low. (Tr. Vol. 1, p. 165; *see* P 118.)

[12] 229. The Court finds that the obstacle course portion of the PAT, if administered as a separate test, would possess some unquantified degree of content validity. Nevertheless, this finding is insufficient to establish the PAT's validity, or job-relatedness, for the reasons set forth below.

b. *The push-ups and sit-ups components of the PAT*

[13] 230. All parties seemingly agree, and this Court finds as a matter of fact, that the push-ups and sit-ups components of the PAT lack content validity. (*See* Tr. Vol. 2, p. 78.)

i. criterion-related validity

[14] 231. The Court also finds that the push-ups and sit-ups components of the PAT lack criterion-related validity.

232. As the parties have stipulated, a test has criterion-related validity to the extent that performance on the test is statistically related to performance on the job. A criterion-related validity study should consist of empirical data showing that the test is predictive of, or significantly correlated with, important elements of job performance. Criterion-related validity is established when an employer shows that scores on its test (even scores as basic as “pass v. fail”) relate in a meaningful way with some measure of job performance (i.e., a “criterion”). (S 62.)

233. The Court credits Dr. Jones' testimony that, to establish criterion-related validity one collects data, consisting of both test (predictor) scores and job performance (criterion) measures, and then performs statistical analyses on the data to show that there is a relationship between the predictor and the criterion such that individuals who have higher test scores tend to have higher levels of performance and individuals who have lower test scores tend to have lower levels of performance. (Tr. Vol. 2, pp. 119–123; Ex. Q–2 and Q–3.)

234. The Court further credits Dr. Jones' testimony that it is unacceptable under professional standards for an individual to examine a test and declare, based on his/her own professional judgment, that it has criterion-related validity when there ***559** is no data to support the judgment. (Tr. Vol. 2, p. 123.)

235. Dr. Davis did not conduct a criterion-related study to attempt to relate performance on the push-ups or sit-ups components of the PAT to police officer performance. (S 77.)

236. With respect to the sit-up component of the PAT, there is no criterion-related study of record that has been identified by the Defendant or its expert as having relevance to this case. (*See* Tr. Vol. 3, p. 148.)

237. With respect to the push-ups component of the PAT, the only criterion-related study involving law enforcement officers that Dr. Davis has identified is his 1988 Anne Arundel

County study. (See Davis Depo. (11/17/04–11/18/04) at pp. 23–24, 26–27, 106–07, 109–110; Ex. 17 at Appendix C (“Relevant Publications and Citations”); *see also* Tr. Vol. 3, p. 148.)

238. At trial, however, counsel for the City, when questioned by the Court, seemed to indicate that—while the City was not completely withdrawing its reliance on the study—the City had chosen to not highlight the Anne Arundel study by having Dr. Davis not testify about it during his direct examination. (Tr. Vol. 3, p. 157.)

239. Regardless whether or not the City withdrew its reliance on the Anne Arundel report, the City did not introduce the report into evidence, and Dr. Davis' reports, Tr. Ex. 16 and 17, do not discuss the Anne Arundel study. Accordingly, the only evidence presented at trial regarding any criterion-related study of a push-ups test involving law enforcement officers was the unrebutted testimony and reports of the United States' experts regarding the Anne Arundel study.¹⁴

240. The Court credits the testimony of Dr. McArdle that one cannot infer any similarities between the Anne Arundel study and the PAT because the two tests are materially different from one another and do not test the same construct. (Tr. Vol. 3, pp. 88–89, Ex. W, pp. 7–8.) As Dr. McArdle credibly explained, the Anne Arundel push-ups test required that the subject perform as many push-ups as possible to exhaustion, without any time limit, thereby measuring the subject's muscular endurance. For example, an individual who performed 18 push-ups in 60 seconds would have received a higher score than an individual who performed 17 push-ups in 15 seconds. (Tr. Vol. 3, pp. 82–83, Ex. W, p. 7.)

241. By contrast, the PAT's push-up component was tied to performance within a short time period and thus (unlike the Anne Arundel study) depended greatly on speed. An individual who could perform 40 push-ups within one minute but could not do push-ups quickly enough to complete 17 push-ups in 15 seconds would have scored well on the Anne Arundel test, but would likely score poorly on the PAT. (Ex. W, p. 7.)

242. Although speed mattered a great deal in terms of performance on the push-ups component of the PAT, nothing in the Anne Arundel report established a relationship between push-up speed and job performance. Accordingly, even if the Anne Arundel study had established that the Anne Arundel push-ups test had some level of criterion-related validity, that

would not establish that the push-ups component *560 of the PAT had criterion-related validity.

243. Moreover, Dr. Siskin credibly testified that, to the extent the Anne Arundel study indicated that the Anne Arundel push-ups test predicted anything at all, the study illustrated that the Anne Arundel push-ups test was unfair to women. (Tr. Vol. 3, pp. 166–67; *id.* at 205; Ex. S at ¶¶ 34–36.)

244. Using Trial Exhibit T–7, Dr. Siskin explained that a statistical analysis contained in the Anne Arundel report showed that both male and female subjects in the study who successfully climbed a six-foot wall did significantly more push-ups than subjects of the same gender who did not successfully climb the wall. Stated otherwise, according to Dr. Siskin, the Anne Arundel push-ups test predicted the subject's ability to successfully climb a six-foot wall. (Tr. Vol. 3, pp. 161–63, Ex. T–7, Ex. S at ¶¶ 35–36.)

245. However, using Trial Exhibit T–8 to illustrate the point graphically, Dr. Siskin explained that the same number of push-ups done by males and females did *not* predict the same likelihood of success at climbing the wall. Rather, females who successfully climbed the wall did approximately the same number of push-ups as males who failed to climb the wall and did many fewer push-ups than males who successfully climbed the wall. Thus, females who performed 21 push-ups were as likely as males who did 33 push-ups to climb the wall successfully. (Exs. T–8, T–7, Tr. Vol. 3, pp. 163–65, Ex. S at ¶ 36.)

246. Dr. Siskin credibly explained that this is an example of “differential validity.” In other words, according to the Anne Arundel analysis, if a man and a woman obtained the same score on the push-ups test, the woman's predicted job performance would be better than the man's. Thus, in order to select men and women who would perform equally well on the job, the employer would have to use different cutoff scores on the push-ups test for men and women. Specifically, the Anne Arundel report indicates that, if men were required to do 33 push-ups, then women should be required to do only 21 push-ups in order to select men and women with an *equal* likelihood of climbing the wall. (Tr. Vol. 3, pp. 164, 20, 166–67, 168.)

247. The Court accepts as credible Dr. Siskin's explanation that, in this circumstance, requiring that men and women complete different numbers of push-ups to pass the test is not “gender-norming,” and it is not using “different standards”

for males and females. Rather it is using the *same* standard in terms of predicted success on the job task at issue (*to wit*, wall climb). (Tr. Vol.3, pp. 166, 168–69.)

248. Based on the unrebutted evidence presented at trial, the Anne Arundel study provides no support for Dr. Davis' opinion that the push-ups component of the Erie PAT has criterion-related validity and, instead, indicates that the push-ups test used in Anne Arundel was unfair to women and was not valid as a stand-alone test.

249. Dr. Davis did not conduct a criterion-related study to relate the push-ups or sit-ups component of the Erie PAT to police officer job performance. (S 77.)

250. Dr. Davis has never validated, used or recommended a test of 15–seconds duration as a measure of muscular endurance. (S 78.)

251. Dr. Davis knows of no other law enforcement agency or research study that has used a 15–second test of push-ups or sit-ups to measure muscular endurance. (S 79.)

252. The City has produced no formal study finding a statistically significant correlation between a push-ups test and/or sit-ups test and performance in a struggle or simulated struggle. (S 80.)

***561** 253. In summary, the Court finds that the City failed to establish at trial that the either the sit-ups or push-ups components of the PAT, standing alone, had criterion-related validity.

ii. Construct Validity

254. The parties have stipulated that the construct validity approach requires both: 1) a showing that the test being validated measures a particular identifiable characteristic (or “construct”); and 2) a showing that the construct is related to job performance. To fulfill the second requirement, the user should show empirically that the test validity relates the particular construct measured by the test to the performance of critical or important work behavior(s). This often requires a criterion-related study. (S 63.)

255. Thus, as Dr. Jones testified, using Trial Exhibit Q–4 as an illustration, in order to establish that the push-ups (or sit-ups) component of the PAT has construct validity, the City

would have to prove that the push-ups (or sit-ups) component is statistically related to other measures of the same construct and that the other measures of the construct are statistically related to job performance. Dr. Jones' testimony is consistent with the parties' stipulated definition of construct validity and was not rebutted by the City. (Tr. Vol. 2, pp. 124–26; Ex. Q–4.)

256. Dr. Davis claims that the push-ups and sit-ups components of the PAT measure the construct of muscular endurance and that muscular endurance is required to engage in a physical struggle. (Tr. Vol. 2, p. 84; Davis Depo. Tr. (11/17/04–11/18/04) at pp. 99–101.)

257. However, neither Dr. Davis nor the City presented any evidence of a statistical relationship between the push-ups or sit-ups component of the PAT and any recognized measure of muscular endurance. (Tr. Vol. 2, pp. 140–41; Ex. Q–7.)

258. In fact, the evidence credibly shows that a push-ups or sit-ups test that—like the push-ups and sit-ups components of the PAT—emphasizes performing a rapid number of repetitions in a brief period like 15 seconds does not measure muscular endurance. (Tr. Vol. 3, p. 116; Ex. W, p. 3 at ¶ 4; *id.* at pp. 7–8; Tr. Vol. 2, pp. 85–86, Davis Depo. Tr. (11/17/04–11/18/04) at p. 112; Tr. Vol. 3 at pp. 79, 83–85, 157.)

259. Indeed, the City stipulated that Dr. Davis has never validated, used or recommended a test of 15 seconds duration as a measure of muscular endurance and that Dr. Davis knows of no other law enforcement agency or research study that has used a 15–second test of push-ups or sit-ups to measure muscular endurance. (S 78–79.)

[15] 260. Because the City has not established that the push-ups component or sit-ups component of the PAT measures the construct of muscular endurance, the City has not established that either component, as a stand-alone test, has construct validity.

261. Moreover, even if the City had established that the push-ups and sit-ups components of the PAT measure muscular endurance, the City presented no evidence of a statistical relationship between muscular endurance and job performance.

262. Although the City added the push-ups and sit-ups components to the obstacle course based on the belief that they would relate to performance in a struggle, the City stipulated prior to trial that it had produced no study

finding a statistically significant correlation between a test of muscular endurance as measured by push-ups or sit-ups and performance in a struggle or simulated struggle. (S 28, 80.)

263. In fact, according to Dr. Davis, success in a struggle depends more on ***562** strength and lean body mass than muscular endurance, and lean body mass correlates *negatively* with muscular endurance. In other words, individuals who have more lean body mass are likely to perform best in a struggle, but worst on a push-ups and sit-ups test that measures muscular endurance. (See Davis Depo. Tr. (11/17/04–11/18/04) at p. 105, 186–88, Ex. W, pp. 8–9; Tr. Vol. 3, p. 105; Tr. Vol. 2, pp. 68–69.)

264. Moreover, to the extent that Dr. Davis' opinion regarding construct validity is based on an assumption that the ability to do more push-ups and sit-ups will mean better performance in a struggle because the same muscle groups used to perform push-ups and sit-ups are involved in a struggle, the Court declines to credit that opinion. (Tr. Vol. 2, pp. 78–79.)

265. Instead, the Court credits Dr. McArdle's testimony that, pursuant to the well-established principle of "exercise specificity," broad measures of general physical fitness like sit-ups and push-ups do not reflect an individual's ability to perform specific physical tasks even if those tasks require the activation of similar musculature. As a simple example, Dr. McArdle pointed out that the person who lifts the most weights at the gym will not necessarily be the best shot-putter, high jumper, boxer or wrestler. Similarly, the person who performs more push-ups (or sit-ups) is not necessarily the one who will perform best in the physical activities performed by police officers. (Ex. W, p. 9; Tr. Vol. 3, pp. 100–102.)

266. Because the City has established neither: 1) that the push-ups and sit-ups components of the PAT measure muscular endurance; nor 2) a statistical link between a recognized test of muscular endurance and any measure of police officer job performance, the City has not established that either the sit-ups component or the push-ups component of the PAT, as a stand-alone test, would have construct validity.

267. In summary, the City and its expert have produced no evidence establishing through any method that the push-ups and sit-ups components of the PAT, standing alone, are valid. Thus, even if it were appropriate methodology for the City to validate the PAT by demonstrating the validity of each component part separately (an approach which we have expressly disapproved here), the City has failed to make the

requisite showing. For the reasons set forth above, the push-ups and sit-ups portions of the PAT are not valid, even when considered as independent test components.

268. The PAT, therefore, fails to satisfy the requirement of job-relatedness.

5. The City has failed to prove that the PAT's passing score corresponds to the minimum level of the tested skills necessary to successfully perform the job of police officer.

269. The parties have stipulated that, in the field of I/O psychology, professionally recognized methods by which cutoff scores appropriately are set include norm-referenced methods, content-related methods and criterion-related methods. (S 68; *see also* Tr. Vol. 2, pp. 144–45.)

270. As previously noted, in setting the PAT's passing standard, the City did not employ any of the foregoing methods in determining whether the PAT's passing standard of 90 seconds corresponds to the minimum level of ability that is required for successful performance as a police officer. (Tr. Vol. 1, pp. 189–90.)

271. Rather, as noted above, the City has stipulated that its representatives who were involved in deciding to use 90 seconds as the passing standard for the PAT did so because they believed that the City ***563** would be requiring a "medium" or "average" level of physical ability, and that seemed "fair" and "the best way to go." (S 44.)

272. Dr. Davis did not do an analysis of the PAT's passing standard using a norm-referenced, content or criterion-related method. (S 82.)

273. In his July, 20004 report, Dr. Davis described three additional methods for setting or validating cutoff scores: a "pacing" (or "concordant") method, a "research design" method, and a "perpetrator" method. (S 69–72.)

274. Dr. Davis did not conduct a "research design" study, or do a "perpetrator" analysis, or use the "pacing" method to justify the PAT cutoff score. (S 75, 76, 81.)

275. Dr. Davis did not do an analysis of the PAT passing standard using the "pacing" or "concordance" method. (S 81)

276. The Erie police officers' responses to Dr. Davis' "job task analysis" survey provide no information about how quickly the officers performed the tasks referred to on the survey

form. The survey responses, therefore, provide no basis for a conclusion that the PAT's 90-second cutoff time is consistent with business necessity. (Ex. 17, App. C, p. 11.)

277. In fact, Dr. Davis never mentioned the 90-second cutoff time on the PAT in either of his expert reports. (*See generally* Exs. 16 and 17.)

278. In short, Dr. Davis did no analysis establishing that the PAT passing standard corresponds to the minimum level of the tested skills necessary to perform the job of police officer successfully.

279. Indeed, at times during the trial it was unclear what Dr. Davis' opinion actually is with regard to whether the 90-second passing time corresponds to the minimum level of physical ability necessary to perform the police officer job successfully. In his July 14, 2004 report, Dr. Davis writes, “[T]he City is at great risk of accepting individuals who cannot perform the rigors of the job since the test's action limits are well below those required on the job,” thus suggesting that the PAT is, in fact, too easy. (Ex. 17, p. 7.) During his deposition in this case, he admittedly stated that “[i]f the test is so incredibly insensitive to allow an 11-year-old girl to pass it, it hardly rises to the definition of a valid test.” (Tr. Vol. 2, p. 91 (quoting 11/18/04 Depo. Tr., p. 184, L 21.)) He also admittedly stated during deposition that “the sensitivity that we have with this fairly minimal test in my opinion really doesn't rise to the level of having a strong predictive power because of the fact the test is so easy.” (Tr. Vol. 2, p. 88 (quoting Depo. Tr. p. 113, L 11.)) And at trial he testified that, even if a more stringent passing standard (e.g., 85, 80 or 75 seconds) were used on the PAT, in his opinion, the PAT still would be “fundamentally defective because it lacks a strength component.” (Tr. Vol.2, p. 93.)

280. Nevertheless, Dr. Davis also stated at trial that, despite his reservations that the PAT is too easy, “it meets categorically the expectation of the minimum, minimum, minimum physical standard” for police officer work. (Tr. Vol.2, pp. 45–46.) Dr. Davis' occasional ambiguity is perhaps best reflected in his response to his own counsel's questioning: when asked whether the PAT “distinguish[es] well for those who pass the test, between those who can do the job and those who can't,” Dr. Davis replied, “Probably not. I clearly believe that it does a very good job of selecting up people who could not do the job, yes.” (Tr. Vol.2, pp. 97–98.)

281. Moreover, Dr. Davis' testimony at times seemed directly at odds with the mandates of *Lanning*. When asked by the Court what the cutoff time should be on this test “if it was, in [his] opinion, to be a *564 valid test within the meaning of *Lanning*,” Dr. Davis initially testified that he is “really quite pleased with” the 90-second cutoff. (Tr. Vol.2, p. 93.) Yet, in his rebuttal testimony, Dr. Davis acknowledged that his “litmus test” for hiring is whether a police officer candidate will *exceed* by some reasonable amount the minimum necessary qualifications to do the job.¹⁵

282. To the extent that Dr. Davis' testimony can be understood as stating that it is his opinion that the 90-second passing standard corresponds to a point at or below the minimum level of the physical abilities necessary to successfully perform the job of police officer, the Court declines to credit that opinion because Dr. Davis provided no reliable basis for such an opinion.

283. Thus, Dr. Davis' testimony does not establish that the 90-second passing standard used by the City corresponds to the minimum level necessary to perform the entry level police officer job successfully.

[16] 284. The Court further finds that the 1994 standard-setting exercise performed by the 19 incumbent police officers does not establish that the PAT's 90-second passing score corresponds to the minimum level necessary to successfully perform the job. On the contrary, the Court finds that the design of the standard-setting exercise made it likely that the passing standard was set too high.

285. The United States' experts, Drs. Siskin and Jones, credibly testified that the use of volunteers in the standard-setting exercise may well have inserted a “bias” into the results because research shows that volunteers tend to be better-than-average performers. (Tr. Vol. 3, pp. 179–80, 182–83; Ex. S at ¶ 19; Tr. Vol. 2, pp. 152–54, Ex. P at ¶ 129; Tr. Vol. 3 at pp. 193, 196–97.) In fact, while Dr. Davis defends the City's use of incumbent volunteer officers in the 1994 standard setting exercise, he acknowledged that it would be a “reasonable basis for a challenge” to assert that the use of volunteers “stacked the deck because these are self-selective people.” (Tr. Vol.2, p. 47.)

286. Dr. Siskin credibly testified that the over-representation of (8) SWAT team members among the 19 volunteers (at a time when the SWAT team constituted no more than 15% of

the City's police force, *see* fn. 3, *supra*)—and the evidence that the non-SWAT volunteers performed comparably or nearly as well as the SWAT members¹⁶—suggests that, as the research *565 would indicate, the deck was indeed “stacked.” (Tr. Vol. 3, pp. 182–83; Ex. S at ¶ 19.)

287. The City and its expert have suggested that the 19 incumbents may not have put forth their best effort in the standard-setting exercise because of a lack of motivation. However, there is no evidence of record to suggest that the incumbents did not put forth their fullest efforts. On the contrary, the only two incumbent witnesses who were questioned on the matter confirmed that they gave their best efforts during the exercise. (Ex. GG; Tr. Vol. 1, pp. 137, 139 (Fetterman testifying); *id.* at pp. 26–27 (Kemling testifying); Kemling Depo. Tr. (12/2/04) at pp. 17–18; *see also* D 131.)

288. Moreover, Dr. Siskin credibly opined that, for the same reasons that volunteer test-takers tend to be better-than-average performers, volunteers also tend to put forth their best efforts when testing. (See Ex. S, ¶ 19.)

289. In addition, the City stipulated that the police department requested volunteers for the standard-setting exercise by distributing a memorandum that informed potential volunteers that the standard-setting exercise was being conducted to set the physical ability standard for new hires. (S 37.) Absent any evidence that the volunteers in fact were not motivated, it is fair to infer that the fact that the exercise was being used to set standards for new police officers would itself serve as motivation for the 19 incumbents to perform well. (See, *e.g.*, Kemling Depo. (12/2/04) at p. 18.) This inference is particularly reasonable in light of the testimony received from several City police officers which generally confirms that incumbent officers rely heavily upon one another in the course of normal patrol duties and they therefore support high fitness standards as a means of ensuring that their peers will be reliable allies in pressure situations. (See generally Tr. Vol. 1, pp. 42–43 (Raszkowski testimony); *id.* at p. 48 (Kress testimony); *id.* at pp. 79, 88 (Kuhn testimony); *id.* at pp. 124–25 (Mangan testimony); *id.* at p. 148 (Fetterman testimony)).

290. Accordingly, the Court credits Dr. Siskin's opinion that the average of the group of 19 volunteers who participated in the standard-setting exercise likely was higher than the average of a representative sample of Erie police officers (or the average of Erie's police officers as a whole) would have been. (Tr. Vol. 3, pp. 180, 182; 175–76; 193–96; Ex. T–14.)

291. Furthermore, even if the 19 volunteers who participated in the standard-setting exercise had been representative, setting a cutoff at the average of a group implies that about half of the group will fail—i.e., if the cutoff corresponds to the minimum level of ability necessary to perform the job successfully, then presumably about half the group cannot do the job. (Tr. Vol. 3, pp. 179, 188; Tr. Vol. 2, pp. 145–46, 149–50; Tr. Vol. 3, p. 31.) Yet the City stipulated that all 19 volunteers were performing their jobs adequately at the time of the 1994 standard setting exercise. (S 39.) By definition, the *average* level of ability of a group of police officers, all of whom have at least the minimum level of ability necessary to perform a job, is not the same as the *minimum* level of skills necessary to perform the job.

292. Dr. Davis attempted to justify the use of incumbent volunteers and the use of their mean scores by explaining that there is an inevitable age-related decline in officer performance following the first day of *566 hire. Thus, it was Dr. Davis' suggestion that no score *less* than the mean could possibly reflect the minimum requirements of the entry level police officer job. In this regard, however, the Court credits the opinion of Dr. McArdle that lifestyle is a more relevant factor than age in terms of predicting the fitness level of veteran police officers, particularly in the age range reflected by the 19 incumbent volunteers. The Court further credits Dr. McArdle's opinion that experience and training can improve officer performance in the essential physical tasks of the job. Therefore, the Court declines to credit Dr. Davis' justification for the use of the incumbent volunteers and their mean scores in setting the PAT cutoff score.¹⁷

293. Finally, the Court finds that the results of the 1994 standard-setting exercise further suggest that the passing standard used by the City were likely set too high. According to Tr. Ex. GG, the results of the standard-setting exercise, only four (or 21%) of the 19 volunteers completed the obstacle course/run in 60 seconds *and* completed 17 pushups in 15 seconds *and* nine sit-ups in 15 seconds. If the 19 volunteers were representative of Erie officers, and one accepted the City's contention that the components of the PAT should be viewed as separate tests, this would imply that 79% of all Erie police officers were not performing their jobs adequately. (Ex. T–11; Tr. Vol. 3, pp. 187–88; Ex. S at ¶ 15; Tr. Vol. 2, pp. 150–51.)

294. However, as indicated previously, the standard-setting exercise was not the same as the PAT actually administered

to applicants, because the actual PAT was given as one test with one overall cutoff time. Dr. Siskin therefore performed a “conservative” calculation of the percentage of the 19 incumbents who, based on the standard-setting data, would be expected to fail the PAT as it was actually administered to applicants. That calculation, which the Court accepts as credible, indicates that approximately 26% of the 19 incumbent volunteers (or 5 of the 19) would have failed the PAT as it was administered to applicants. (Tr. Vol. 3, pp. 188–90; Ex. T-11.)

295. Based on all the available information, including the volunteer sample and the conservative nature of his calculation, Dr. Siskin estimated that the true incumbent failure rate on the PAT would be at least a quarter and maybe as high as one third to one half. (Tr. Vol. 3, p. 190.) The Court accepts this estimation as credible.

296. The City presented no evidence that any substantial portion—a half, a third or a quarter—of the City's police officers could not perform their job successfully due to an inadequate level of the physical skills measured by the PAT.

297. Certain facts relative to the female contingent of the Erie police force provide additional evidence that the PAT's passing standard was likely set too high. For example, we note that the majority (five) of the eight female officers currently on the City's police force were hired before the City began using the PAT in 1994. Two of the five females, Detective Kemling and Detective Sergeant Mangan, were hired after passing the physical agility test the City used to screen applicants in 1992, which was a gender-normed test that required female applicants to complete only *567 four push-ups in 40 seconds and 12 sit-ups in 60 seconds. (Tr. Vol. 1, pp. 30–31; Ex. II; Tr. Vol. 1, pp. 126–27; Ex. H; Mangan Depo. Tr. (12/3/04) at pp. 5, 9; Depo. Ex. 96; Ex. AA & BB, Request No. 41.)

298. The City has presented no evidence that Detective Kemling or Detective Sergeant Mangan were ever unable to perform the physical tasks required by their jobs. On the contrary, Detective Sergeant Mangan has been promoted twice and received a commendation for her police work. (Tr. Vol. 1, p. 127.) Detective Kemling has been a field training officer and has been promoted to the rank of detective. (Kemling Depo Tr. (12/2/04) p. 6, Tr. Vol. 1, p. 19.)

299. Officers Raszkowski, Szoszorek and Szocki, the three of the City's eight female officers who were hired during the period when the City was using the PAT, barely passed it.

300. Officer Szocki took the PAT in 1994 and passed with a score of 90 seconds (i.e., with no time to spare). (Tr. Vol. 1, pp. 98–99, 198–99; Ex. L; Szocki Depo. Tr. (12/3/04) at p. 10; Depo Ex. 102; *see also* D 124.)

301. Officer Szoszorek took the PAT in 2002, the year in which the City changed the passing standard to 95 seconds due to inclement weather. She passed the PAT with a time of 93 seconds (i.e., with two seconds to spare). (Tr. Vol. 1, pp. 104–105; Ex. M; Szoszorek Depo Tr. (12/3/04) at pp. 10–11; Depo. Ex. 101; *see also* D 125.)

302. Officer Raszkowski was a Pennsylvania high school state champion in the two-mile run and was running four to six miles and doing 20 push-ups and about 100 sit-ups three to four times per week when she applied to be an Erie police officer. Officer Raszkowski failed the PAT in 1996 when she inadvertently discontinued her sit-ups, believing she had completed that portion after performing only seven sit-ups. After being advised to keep going, Ms. Raszkowski resumed her sit-ups, but failed the PAT by one second. She retook the PAT in 1998 and passed with a score of 87 seconds (i.e., with 3 seconds to spare). (Tr. Vol. 1, pp. 33–34; 39–42; Ex. J; Raszkowski Depo. Tr. (12/3/04) at pp. 6–7, 9, 12–13, 23–24; *see also* D 123.)

303. The City offered no evidence that any of the three female officers hired after barely passing the PAT was a minimally-performing police officer or had only the minimum level of physical ability necessary to perform the job. In fact, Officer Raszkowski, who attended the Mercyhurst College Municipal Police Training Academy, won an award for outstanding physical conditioning at the Academy. (Tr. Vol. 3, pp. 61–62; Ex. I.)

304. The 1994 standard-setting exercise simply does not establish that the passing standard the City used on the PAT reflects the minimum level of physical ability necessary to perform the job of entry-level police officer successfully and, on the contrary, suggests that the City likely used a passing standard that was too high.

305. In summary, the City has failed to prove that the PAT is job-related for the position in question. Moreover, even if the City had proved that the PAT was job-related, the City failed

to prove that the passing standard it used corresponds to the minimum level of physical ability necessary to successfully perform the police officer job.

II. CONCLUSIONS OF LAW

1. To the extent any of the Court's findings of fact may be considered conclusions of law, such findings are incorporated herein.
2. Jurisdiction over this case is proper under [28 U.S.C. § 1345](#) and [42 U.S.C. § 2000e-6\(b\)](#). Venue is proper in this district under [28 U.S.C. § 1391\(b\)](#).

***568** [17] 3. Under the disparate impact theory of employment discrimination first articulated by the Supreme Court in [Griggs v. Duke Power Co., 401 U.S. 424, 91 S.Ct. 849, 28 L.Ed.2d 158 \(1971\)](#), and later codified by Congress in the Civil Rights Act of 1991, Title VII prohibits employment practices which, although "fair in form," are nevertheless "discriminatory in operation." [Id. at 431, 91 S.Ct. 849.](#)

4. The applicable burdens of proof in a disparate impact case under Title VII are set forth at § 703(k) of the statute, [42 U.S.C. § 2000e-2\(k\)](#). That provision permits law enforcement agencies to assess candidates for police officer positions with a physical test that results in a disparate impact on females only if the test is "job related for the position in question and consistent with business necessity." [42 U.S.C. § 2000e-2\(k\)\(1\)\(A\)\(i\).](#)

[18] 5. Because of the Court's previous ruling and conclusion that the Defendant's use of the PAT had an adverse impact on female candidates for the job of police officer for the City of Erie, the burden at trial was upon the Defendant to prove that their use of the PAT was "job related" and "consistent with business necessity." *Id.* This involves a 2-step inquiry in the Third Circuit, corresponding to each requirement. See [Lanning v. SEPTA, 181 F.3d 478, 489 \(3d Cir.1999\)](#) ("*Lanning I*") ("Congress chose the terms 'job related for the position in question' and 'consistent with business necessity.'") (emphasis in the original).

[19] 6. As required by the opinions of the Court of Appeals for the Third Circuit in [Lanning I, 181 F.3d at 481](#), and

[Lanning v. SEPTA, 308 F.3d 286 \(3d Cir.2002\)](#) ("*Lanning II*"), the standard for assessing the "business necessity" of a discriminatory passing standard or score on an entry-level police officer test is whether the passing standard "reflects the minimum qualifications necessary to perform successfully the job in question." [Lanning I, 181 F.3d at 489](#); *see also* [United States v. State of Delaware, 2004 WL 609331, *1 \(D.Del., March 22, 2004\)](#).

7. Contrary to the City's assertion, nothing in the Third Circuit's opinions indicates that there is an exception to this business necessity standard when an employer asserts that it is trying to "improve" its workforce.

A. Job Relatedness

8. Given the City's concession, and this Court's findings, that the City's use of the PAT as a unitary test caused a disparate impact against female applicants, the use of the PAT as one test with one passing standard is the challenged "employment practice" at issue for which the City had to prove job-relatedness and consistency with business necessity.

[20] 9. For the reasons outlined in more detail above, the City failed to prove by a preponderance of the evidence the job-relatedness (or validity) of the PAT as a unitary test—that is, as it actually was used by the City.

10. Moreover, even if it were acceptable for the City to demonstrate the validity of each of the individual components of the PAT as if each were a separate, stand-alone test, the City has not made such a demonstration by a preponderance of the evidence.

11. More specifically, the City has not proved the validity of the push-ups or sit-ups component of the PAT under the validation strategies generally followed by the employment testing profession ¹⁸ or by any ***569** of the other methods that the City's expert claimed could be used to validate a physical test.

12. On the contrary, to the extent that evidence regarding the validity of the push-ups component was presented to the Court, the evidence suggests that the push-ups component is not valid. As to the sit-ups component, there was a complete absence of proof by the City as to its validity.

13. Because the City failed to meet its burden of proof as to the “job-relatedness” requirement under Title VII, the City’s use of the PAT between 1996 and 2002 violated that statute. Accordingly, the Court technically need not reach the issue of whether the passing standard on the PAT was set at a level corresponding to “the minimum qualifications necessary to perform successfully the job in question.”  *Lanning I*, 181 F.3d at 489. We nevertheless do so for the sake of judicial expediency.

B. Consistency with Business Necessity

[21] 14. Even if the City had established “job-relatedness” with respect to the PAT, the City has not established that its use of the PAT with the 90-second cut-off score was consistent with business necessity in that it has not proved by a preponderance of the evidence that the PAT’s passing standard corresponds to the minimum qualifications necessary to successfully perform the job of police officer.

15. On the contrary, as discussed more fully above, there is substantial evidence of record which suggests that the cut-off score for the PAT was likely set too high.

16. In *Lanning II*, the Third Circuit explained that an employer could show that a passing standard reflects the “minimum qualifications necessary” for successful performance by showing that individuals who pass the test are “likely to be able to do the job” whereas individuals who fail the test “will be much less likely to successfully execute critical policing tasks.”  *Id.*, 308 F.3d at 291. The City has not presented any evidence to make such a showing.

17. By the City’s own admission, its expert performed no analysis of the PAT’s passing standard. His bare opinion is insufficient to establish that the passing standard used by the City corresponds to the minimum qualifications necessary for successful performance.  *Lanning I*, 181 F.3d at 491–92 (“judgment alone is insufficient to validate an employer’s discriminatory practices”).

*570 18. Moreover, to the extent Dr. Davis performed any “study” at all, the Court is mindful of the Third Circuit’s admonition that “studies done in anticipation of litigation to validate discriminatory employment tests that have already been given must be examined with great care due to the danger of lack of objectivity.”  *Lanning I*, 181 F.3d at 481 (citation omitted).

19. Finally, to the extent Dr. Davis opined that “more is better” with respect to the physical capabilities of police officer candidates, we note the Third Circuit’s observation that “[i]t is unlikely that such a [‘more is better’] study could validate rankhiring with a discriminatory impact based upon physical attributes in complex jobs such as that of police officer in which qualities such as intelligence, judgment, and experience surely play a critical role.”  181 F.3d at 493 n. 23.

20. The City’s 1994 standard-setting exercise does not demonstrate that the PAT’s passing standard is consistent with business necessity. While this Court does not question the City’s good faith in conducting the standard-setting exercise, the decisions made by the City in administering that exercise and in choosing the cut-off score for new officer applicants made it probable that the PAT’s passing standard would be set at an inappropriately high level, both because the City used a non-representative sample of 19 volunteers and because the City chose to utilize the average scores of the volunteers, all of whom the City admitted were performing their jobs at least adequately, rather than determining the level which distinguished successful from unsuccessful performers.

21. The fact that the City’s 1994 standard-setting exercise resulted in a passing standard that would exclude substantial numbers of incumbents who were performing the job successfully suggests that the passing standard did not correspond to the minimum qualifications necessary to perform the job successfully. See  *Lanning I*, 181 F.3d at 494 n. 24 (evidence that incumbent officers failed the physical fitness test yet successfully performed the job tends to show that the passing standard does not correlate with the minimum qualifications necessary to perform successfully); *United States v. State of Delaware*, 2004 WL 609331, *20 (passing standard that would eliminate a substantial number of individuals who would perform the job at an acceptable level does not correspond with the minimum skill level necessary to do the job).

22. The City cannot justify a passing standard set at the average of a sample of officers, all of whom were performing the job successfully and all of whom volunteered to participate in the standard-setting exercise, simply by pointing out that, on average, the officers in the sample were older than the typical applicant for an entry-level police officer job in the City of Erie. The ages of successfully

performing officers, without more, cannot establish that their average performance corresponds to the minimum qualifications necessary to perform the job successfully.

C. The City's Liability

[22] 23. Because the City has failed to prove, by a preponderance of the evidence, that its use of the PAT was both job-related and consistent with business necessity, the City has failed to rebut the presumption that its use of the PAT constituted an "unnecessary barrier" to employment.

 *Griggs*, 401 U.S. at 431, 91 S.Ct. 849.

24. By using the PAT as a screening device for the selection of entry-level police officers, the City of Erie violated Title VII between 1996 and the date (during the course of this litigation) on which the City ceased using the last eligibility list resulting ***571** from any selection process in which the PAT was administered.

25. The nature of the specific relief to be awarded and the entitlement of particular individuals to relief will be determined in the relief phase of this case, following entry of judgment in favor of the United States with respect to liability.

III. CONCLUSION

In rendering these findings of fact, the Court wishes to stress again that it does not question the City's good faith in developing and utilizing the PAT. On the contrary, the Court acknowledges that the City made genuine efforts to develop a physical agility test that would responsive to the Bureau's needs, all the while limited by budgetary concerns and, at times, by objections raised by the police officers' union. The

Court further acknowledges that the City took steps to make its hiring process more amenable to females by modifying its administration of the PAT over the years.

The Court is also sympathetic to the concerns that various police officers have expressed, quite legitimately, about maintaining the physical integrity of their force, particularly as officers must regularly rely on one another in waging the war on crime. Accordingly, in rendering these findings and conclusions, the Court in no way advocates the watering down of hiring standards in such a fashion as to jeopardize either the lives of the officers or the health and safety of the public. Nor should these findings and conclusions be interpreted as a call for quota systems, which would themselves violate the mandates of Title VII. (See  *Lanning I*, 181 F.3d at 490 n. 15 ("Nothing in the *Griggs* business necessity standard requires employers to hire employees in numbers to reflect the ethnic, racial or gender make-up of the community.").)

Simply stated, we are bound by the dictates of Title VII as interpreted by the Third Circuit in *Lanning I* and *Lanning II*. For all the reasons set forth herein, the City has failed to meet its burden of proof with respect to its use of the PAT during the time period challenged in this litigation. Accordingly, the Plaintiff is entitled to a judgment in its favor with respect to the liability phase of this case.

An appropriate Order of Judgment is being filed simultaneously with these findings of fact and conclusions of law.

All Citations

411 F.Supp.2d 524

Footnotes

¹ References to the trial transcript are indicated by the volume number which corresponds to the day of trial, i.e., Tr. Vol. 1 (3/7/05 [Doc. # 78]), Tr. Vol. 2 (3/8/05 [Doc. # 79]), Tr. Vol. 3 (3/9/05 [Doc. # 80]), Tr. Vol. 4 (3/10/05 [Doc. # 81]). References to trial exhibits are as follows: Tr. Ex. ____ (Plaintiff's trial exhibits are denoted by letters while Defendant's trial exhibits are denoted by numbers.). References to the parties' stipulations [see Doc. # 61] are indicated as: "S ____." References to Plaintiff's Post-Trial Proposed Findings of Fact and Conclusions of Law [Doc. 83 and 85] are designated "P ____." Defendant's proposed findings of fact and

conclusions of law [see Docket # 68] are designated as: "D ____." Excerpts from deposition testimony admitted at trial (see Plaintiff's Deposition Designations) are indicated by "Depo. Tr. (date) at pp. ____."

- 2 At the time the City developed and began using the PAT, the Civil Service Commission was, at least de facto, the final decision-making body regarding what physical agility test would be given to entry-level police officer applicants. (S 1.)
- 3 Defendant's proposed findings of fact and conclusions of law [Docket No. 68] are misnumbered such that numbers 118–121 are used twice for different paragraphs. Our reference is to the second of Defendant's proposed findings number 118, found at unnumbered page 5.
- 4 As originally developed by Lt. Bowers, the PAT obstacle course/run included two events (rather than one) intended to simulate climbing through a window. This was later modified. (S 22.)
- 5 The PAT administered by the City between 1994 and 2002 also included a separate trigger-pull component performed after, and only if, the applicant successfully completed the other three components of the test. Because the trigger-pull event was not part of the timed portion of the PAT, the parties did not consider it to be part of the challenged PAT for purposes of this case.
- 6 In 1994, approximately 10–15% of the City's police officers were members of the SWAT team. (S 38.)
- 7 Chief Bowers testified that, based on his observations, males had more difficulty with sit-ups while females had more difficulty with push-ups. By increasing the number of required sit-ups and decreasing the number of required push-ups, Bowers felt the test would be more equitable toward women. (Tr. Vol.1, pp. 178–180.) The net effect of this change, in Chief Bowers' view, was to reduce the importance of upper body strength in passing the PAT.
- 8 The City of Erie has since discontinued its use of the PAT challenged by the United States in this lawsuit. (S 48.) Instead, the City will require that applicants for entry-level police officer positions already be certified as law enforcement officers by the Commonwealth of Pennsylvania (i.e., obtain Act 120 certification). In order to enter the police academy, an individual must take the physical agility test mandated by Pennsylvania's Municipal Police Officer Education and Training Commission (MPOETC) and perform each of the components of the test at the 30th percentile of the individual's gender and age group. In order to complete academy training and obtain Act 120 certification, an individual must take the same MPOETC test and perform each of the components of the test at the 50th percentile of the individual's gender and age group. Before hiring any applicant listed on an eligibility list, the City will henceforth require that the applicant retake the MPOETC test and score at the 30% percentile. (See P 46, 47.)
- 9 Consistent with the practice of the parties and their expert witnesses at trial, the term "validity," when used in these findings, means job-relatedness, and "valid means job-related."
- 10 Nevertheless, at trial Dr. Davis testified that the City's decision to place the push-ups/sit-ups components at the beginning of the PAT in 2002 is of little moment and that, whether they were at the beginning or end, "the concept is essentially the same." (Tr. Vol.2, p. 61.)
- 11 Females comprised 22% of the total. (Ex. 17, p. 9.)
- 12 Dr. McArdle explained that, because of physiological difference between the sexes, women have to work harder in order to perform the same number of push-ups as their male counterparts. Consequently, if a woman and man both run the obstacle course in 60 seconds and then complete 17 push-ups, it can be inferred that the male was more fatigued by the run than the female. (Tr. Vol.III, p. 86.)

13 Conversely, in years when the push-up/ sit-up components preceded the obstacle course run, the amount of time taken to complete the requisite push-ups and sit-ups would affect the amount of time the candidate would have to complete the obstacle course run.

14 The Anne Arundel report provides no support for an opinion that the sit-ups component of the PAT or any other sit-ups test has criterion-related validity for the simple reason that the report said nothing about sit-ups. (See Ex. W, p. 8 (noting that the Anne Arundel study says nothing about sit-ups).)

15 In defending the City's use of incumbent volunteers (presumably, a highly motivated group of individuals) in the 1994 standard setting exercise, Dr. Davis testified as follows:

... If you're trying right now to take a snapshot in time that is supposed to go forth from 20 to 30 years of employment, that's a long way to look over the horizon. The best probability of getting the person that is going to go the distance is trying to find the person who does exceed by some reasonable margin the minimal levels of what the job is. And certainly I'd like to have at least the average of what I consider to be a reasonably motivated cohort.

* * *

THE COURT: I just want to make sure I caught something. What you say you're looking for are people who will exceed by some reasonable amount the minimum necessary qualifications to do the job, that's what you just said, isn't it?

[DR. DAVIS]: You're right, your Honor. It's what I call the litmus test. It's not that we want to find the tensile breaking strength of a rivet for the wing of an airplane, we would like to have a safety factor above that....

(Tr. Vol.4, pp. 23–24.)

16 The City asserts that Dr. Siskin's testimony is rebutted by the fact that the average performance of the group of incumbents who participated in the standard-setting exercise would not change substantially if the 8 SWAT team members were excluded. We find, however, that—on the contrary—this evidence buttresses the conclusion *not* that the SWAT team members were themselves low or average performers but, rather, that the group as a whole, including non-SWAT team incumbents, were better-than-average performers. (See D 95, D 96; Tr. Vol. 2, p. 48; *id.* at 164–65; Tr. Vol. 3, pp. 182–86.)

17 The Court acknowledges that the mean time score of the 19 incumbent volunteers was actually 87 seconds. Thus, police officer candidates received the benefit of a 3-second cushion when the passing score was raised to 90 seconds. However, the Court continues to find, notwithstanding this 3 second benefit, that the use of the mean scores of the 19 volunteers tended to artificially inflate the passing score beyond the level reflective of the minimum physical requirements of the job.

18 At trial, Dr. Davis made the assertion that the Third Circuit “threw out the entire SIOP principles” in *Lanning I*. (See Tr. Vol. 4, p. 6.) We assume Dr. Davis was referring to footnote 20 of the court's opinion, in which it stated that “[t]o the extent that the [Principles for the Validation and Use of Personnel Selection Procedures ('SIOP Principles')] are inconsistent with the mission of *Griggs* and the business necessity standard adopted by the Act, they are not instructive.”  181 F.3d at 493 n. 20. As is self-evident from the court's chosen language, the court did not “throw out” or otherwise invalidate the SIOP Principles in their entirety. Rather, the court's comment was made in reference to the fact that the district court in *Lanning I* had improperly failed to consider whether Dr. Davis' suggested cut-off score for SEPTA transit police officers reflected the minimum aerobic capacity necessary to successfully perform that job. The Third Circuit was criticizing the district court for having adopted Dr. Davis' suggested cut-off score as “readily justifiable”—a standard which appeared to be derived from the SIOP Principles—rather than analyzing whether the suggested cut-off score

was truly reflective of the minimum job requirements as *Griggs* mandates. Accordingly, we see nothing in the *Lanning I* decision which requires us to abandon, wholesale, any reliance on the *Principles* and *Standards* discussed herein. Nor do we consider our reliance on those *Principles* and *Standards* in the context of this case to be in any way “inconsistent with the mission of *Griggs* and the business necessity standard adopted by the Act.”  181 F.3d at 493 n. 20.

End of Document

© 2025 Thomson Reuters. No claim to original U.S. Government Works.