UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF ILLINOIS
EASTERN DIVISION

ARTHUR L. LEWIS, JR., et al.,	)	
	)	
Plaintiffs,	)	
	)	Judge Joan B. Gottschall
v.	)	
	)	Case No. 98 C 5596
CITY OF CHICAGO,	)	
Defendant.	)	

MEMORANDUM OPINION AND ORDER

Plaintiffs, the African-American Fire Fighters League of Chicago (the "League") and a class of African-Americans who applied for entry-level firefighter jobs with the Chicago Fire Department ("CFD") and who scored between 65 and 88 on an entrance exam administered to firefighter candidates in 1995 (the "1995 Test") by defendant City of Chicago ("City"), have sued the City alleging violations of Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e, *et seq*. Specifically, plaintiffs argue that the City's decision to select only those firefighter applicants who scored at least 89 points on the 1995 Test had an unjustified adverse impact on African-American applicants. 42 U.S.C. § 2000e-2(k)(1)(A)(i). The City concedes that its hiring procedure had an adverse impact on African-American applicants, but argues that: (1) the 1995 Test validly measured some of the cognitive skills necessary to training and for performing the job of firefighter; and (2) the City's decision to set a cut-off score of 89 was justified by administrative convenience in that the City wanted to limit the number of applicants that it accepted for further evaluation.

The court conducted an eight-day bench trial on plaintiffs' claims in January 2004. The parties submitted post-trial findings of fact and conclusions of law, post-trial motions for ruling on

unresolved motions *in limine* and evidentiary objections, motions related to the issue of the League's standing to join plaintiffs' class claims against the City, and supplemental authority related to the court's May 25, 2000 ruling denying the City's motion for summary judgment on grounds of untimeliness. These matters were fully briefed before the court as of August 2, 2004.

For the reasons that follow, the court rules in favor of plaintiffs on the liability aspects of their discrimination claim against the City. The court finds that the City has not carried its burden of proof in this case; it has not proven that its decision to hire only those applicants who scored 89 and above on the 1995 Test was consistent with business necessity. To the contrary, the evidence at trial demonstrated that: (a) the 1995 Test may not be a reliable measure of the four cognitive abilities it was intended to measure; (b) the 89 cut-off score was a statistically meaningless benchmark; (c) even if the 1995 Test could reliably measure what it was supposed to measure, it could not distinguish between those who were qualified for the position of CFD firefighter and those who were not; and (d) less discriminatory, and equally convenient, selection strategies were available. In short, the City has not proven that its discriminatory selection process was justified. The court, therefore, finds the City's selection procedure unlawful under Title VII.

BACKGROUND [1]

The City's Hiring Procedure

Since 1996 and through the present, the City has relied on test scores from a written exam given in 1995 as the primary basis for selecting entry-level firefighters. On July 26 and 27, 1995, the City administered the exam to approximately 26,000 people who satisfied the minimum registration requirements of: (1) being at least 18 years old; (2) living in the City of Chicago; and (3) holding a high school degree or its equivalent. After scoring the exam, the City decided that, with exceptions for military veterans and certain paramedics, only applicants with scores of 89 and higher – out of a possible 100 points – would be eligible to proceed to the next phase of the hiring process, a physical abilities test. Applicants who passed the physical abilities test were subject to a background investigation, and those passing the background check were given a medical exam and a drug test. Once an applicant passed all of the City's preliminary tests, he or she was hired as a candidate firefighter. To become a full firefighter with the CFD, candidates were required to complete the Chicago Fire Academy's (the "Academy's") training program and to pass the Illinois board certification exam.

It is undisputed that the City's decision to set the cut-off score for the 1995 Test at 89 points had a severe disparate impact on African-American applicants. Of the 26,000 applicants taking the

[1] Plaintiffs have moved to admit into evidence several exhibits (Pl. Exs. 16, 18, 37-39, 42, 43-49, 50, and 55-61) that were introduced at trial over the City's objection. The City continues to object to the admission of this evidence, primarily on the grounds of unfair prejudice. Fed. R. Evid. 403. In the context of a bench trial, however, Rule 403 objections have no logical application and are routinely overruled. As this case was not tried before a jury, the court fails to understand how the City will be prejudiced by the court's consideration of any and all material introduced during the bench trial. To the extent the material was relevant and probative of plaintiffs' case or the City's defense, the court has so considered it, and to the extent the evidence was irrelevant or unfairly prejudicial, the court has disregarded it. Plaintiffs' motion for the admission of evidence is granted.

exam, 11,649 (45% of test takers) were white and 9,497 (37%) were African-American. It is undisputed that there is no difference between whites and African-Americans in firefighter performance. However, there were pronounced group differences in performance on the 1995 Test: the difference between the mean score of whites and the mean score of African-Americans on the 1995 Test was almost a full standard deviation.[2] The disparate impact of the 1995 Test was heightened by the City's use of the 89 cut-off score. Approximately 12.6% of whites compared to 2.2% of African-Americans scored 89 or above. In other words, the City's decision to select only those applicants who scored 89 and above meant that white applicants were five times more likely than African-Americans to advance to the next stage of the hiring process.

From 1996 to 2001, the City advanced applicants for entry-level firefighter positions from the "well-qualified" pool (those who scored 89 and above on the 1995 Test). The City made a few exceptions to the 89 cut-off score: from 1996 to 2001, the City hired approximately 182 paramedics and 325 military veterans with scores between 65 and 88. The City considered those paramedics and veterans qualified for the position of probationary firefighter despite their lower scores on the 1995 Test. By the summer of 2001, the City had run out of candidates from the "well-qualified" pool and decided to begin processing applicants at random from the "qualified" pool (those scoring between 65 and 88). Most of the 100 cadets entering the Academy in the fall of 2002 received a score between 65 and 88. That class graduated from the Academy at the end of April 2003. There is no evidence that those firefighters from the Academy class of 2003 are any less qualified, in any respect, than those hired with test scores of 89 or above. Moreover, virtually all candidates who have entered

[2] A standard deviation is a statistical measure of the dispersion of results from the mean. The standard deviation tells us how far a typical member of a population is from the average member of that population.

the Academy – regardless of test performance – have successfully completed their training and passed the state certification examination.

The Development of the 1995 Test

The City devoted considerable resources to creating the 1995 Test. The City hired the consulting firm Human Performance Systems, Inc. ("HPSI"), and Dr. James Outtz, an industrial organizational psychologist with extensive experience designing and evaluating entrance tests, to develop its firefighter exam. The 1995 Test was constructed using a "content-oriented" test validation strategy, which measures whether the content of the test reflects important aspects of performance on the job for which the candidates are being evaluated. The City chose not to pursue a "criterion-related" validation strategy, which uses empirical data to show that the test can predict (or at least correlates to) the test taker's ability to perform the job. The City avoided the "criterion-related" approach because it did not have the data required to link test performance to job performance: the City had security concerns about giving the test to incumbent firefighters, and the CFD does not conduct formal evaluations of firefighter performance.

The City's "content-based" job analysis aimed to: (1) identify the tasks performed by firefighters on the job; (2) identify the knowledges, skills, and abilities required to perform the tasks effectively; (3) eliminate from consideration for testing those tasks that were unimportant or done infrequently and those abilities that were not "needed day one" (*i.e.*, prior to training); and (4) link the remaining knowledges, skills, and abilities to tasks that require them. Those knowledges, skills, and abilities that survived the job analysis procedures were termed "critical" or "essential." The job analysis for Chicago firefighter proceeded in three broad phases: (a) a "job inventory," which identified the tasks and abilities required to perform the job; (b) a "job analysis questionnaire" to

collect ratings from incumbent firefighters of the job tasks, knowledges, skills, and abilities identified by the job inventory; and (c) a "linkage questionnaire" which required incumbent firefighters to link important knowledges, skills, and abilities identified from the job analysis questionnaire to "task groups" comprising the firefighter job.

The job analysis conducted by Dr. Outtz and HPSI yielded a list of 46 skills deemed critical to the job of Chicago firefighter. Of these 46, 18 were deemed "essential" and "needed day one," meaning they were required of firefighter candidates before training at the Academy. Of those 18 "needed day one" abilities, Dr. Outtz and HPSI determined that 8 were physical skills, 3 were essentially untestable because of their intangible qualities, and 7 were "cognitive" skills appropriate for testing on a written exam. Of those 7 cognitive abilities, 4 were tested by the 1995 exam: (1) the ability to comprehend written information; (2) the ability to understand oral instructions; (3) the ability to take notes; and (4) the ability to learn from or understand based on demonstration.

The 1995 Test had two parts, a multiple choice "pencil and paper" section and a video demonstration section. The written portion of the exam was designed to measure an applicant's ability to comprehend written information. The 1995 Test was written at a twelfth-grade reading level, which approximated the reading level of the materials used at the Academy and written CFD policies and procedures. The video portion of the exam was designed to measure an applicant's ability to understand oral instructions, ability to take notes, and ability to learn from or understand based on demonstration. The subject of the video was a fictitious mechanical device called a "fuel converter system." Applicants were first shown the device and its components on the video screen, along with a "trainer" and "trainee" using the device, while an off-camera narrator explained its operation. Applicants were then asked questions about the device based on the information that had

just been shown on the video. Prior to taking the exam, applicants were given reference booklets that contained the written material upon which the test questions would be based and a description of the fictitious device that would be the subject of the video component. Applicants were permitted to refer to these materials during the exam.

The Scoring Of The 1995 Test And The City's Selection Of Candidates

Raw scores on the written and video components of the 1995 Test were: (1) corrected according to standard statistical methods; (2) weighted at 15% and 85%, respectively, to reflect the importance of the cognitive abilities being tested in each section; and (3) converted to a 100-point scale. The distribution of scores ranged from a low score of 12 points to a high score of 98 points with an average score of 75. The City set the passing score for the exam at 65, which was one full standard deviation below the mean. The City concedes that every applicant scoring 65 and higher on the 1995 Test possessed the minimum level of cognitive ability to master the Academy curriculum and perform the job of firefighter. Out of approximately 26,000 people taking the exam, 93.45% of whites and 72.3% of African-Americans "passed" with a score of at least 65 points and were thus considered "qualified" to advance in the hiring process.

With the results of the 1995 Test in hand, the City's Deputy Commissioner of Personnel, Robert Joyce, set a cut-off score of 89, selecting only those applicants who scored at least 89 points for further evaluation. That decision had a profound effect on the racial makeup of the candidate pool. The so-called "highly qualified" pool – those who scored 89 and above – from which the City hired all of its entry-level firefighters from 1996 to 2001, was comprised of approximately 5.4 times more whites than African-Americans. By contrast, the "qualified" pool of applicants – those who passed the 1995 Test by scoring a 65 or above – was comprised of only 1.3 times more whites than

African-Americans. In arriving at the cut score of 89, Joyce testified that the City considered: (1) the hiring needs of the CFD during the three to five years the City planned to rely on the results of the 1995 Test; (2) the fairness to applicants of identifying several thousand applicants as "qualified" for further processing when only several hundred of them would ever be hired; and (3) the adverse impact of setting the cut score at various points higher than the passing score of 65. Joyce also stated that he assumed, based on Dr. Outtz's analysis of the test scores, that the 1995 Test was valid, meaning "you can make some inferences from [the test] scores. The higher scores – in a very general way, higher scores are more predictive of success than lower scores."

However, Joyce's assumption was not correct and his decision to set the cut-off score at 89 did not account for the statistical properties of the 1995 Test. Dr. Outtz testified that, based on his statistical analysis of the 1995 Test, he initially recommended that the City set the cut-off score by counting down from the top score of 98 in 13-point increments. He arrived at his 13-point band by calculating the "standard error of the difference," an index measuring the extent to which a difference in scores is statistically significant or due to chance, based on the internal "reliability" of the 1995 Test. The reliability of a test refers to the extent to which scores are free from random error, *i.e.*, the extent to which retesting of a given applicant is expected to yield a consistent result. Since retesting was not an available option, Dr. Outtz instead calculated reliability by comparing the consistency of answers given to different questions on the 1995 Test by the individual applicants who took it. By Dr. Outtz's calculations, the 1995 Test had a reliability coefficient of .77, meaning that approximately 23% of the variance in individual scores was due to random error.

Based on that calculation, Dr. Outtz determined that there is no statistical difference between any two scores from the 1995 Test that are within 13-points of each other, *i.e.*, a score of 98 cannot

be meaningfully distinguished from a score of 85. Given the statistical properties of the 1995 Test, Dr. Outtz concluded that there was a "psychometric basis" – a basis rooted in cognitive analysis – for setting the cut score using that 13-point band. As he explained, "[T]here is a psychometric basis for saying, for reaching the inference that the people who are within the band that I had determined . . . have more of the abilities measured by the test than people outside the band."[3] Dr. Outtz also testified, however, that there was no psychometric basis for setting the cut score at any point within the 13-point band. In other words, in Dr. Outtz's opinion, a score of 89 could not be statistically distinguished from a score of 87 or 88, two lower scores within the 13-point range below the top score of 98. Because the standard error of the difference was so large, Dr. Outtz discussed with the City the possibility of randomly selecting candidates from the pool of applicants who passed the 1995 Test with a score of 65.

Despite Dr. Outtz's conclusion that the 1995 Test could not distinguish between scores within 13 points of each other, the City decided to set the cut score at 89, only 11 points below the highest score. Joyce testified that he made the decision to hire only those applicants scoring 89 and above: (a) against Dr. Outtz's recommendation: and (b) with full awareness of the 1995 Test's disparate impact on African-Americans generally, and of the even greater disparate impact on African-Americans caused by setting the cut score at 89. Joyce testified that he set the cut-off score at 89 because it was the most administratively convenient way to trim the list of potential applicants to a manageable number while still fulfilling the hiring needs of the CFD.

---

[3] For example, according to Dr. Outtz's testimony, a cut-off score of 85 would be somewhat defensible as it would "capture" all of the scores that are indistinguishable from the top score of 98. Although there would be no way to differentiate candidates within the 13-point range, there would be a basis for claiming that an individual who scored 98 has greater tested skills than an individual who scored 84.

Approximately six months after the 1995 Test was given, the City sent all applicants notices of their final scores. The City grouped the scores into three categories: applicants who scored 89 and above were considered "well qualified" and were eligible to advance in the hiring process; applicants who scored between 65 and 88 were considered "qualified"; and applicants who scored below 65 failed the examination. Applicants in the "qualified" pool – the plaintiff class in this case – were informed that, due to the large number of applicants who received higher scores, and based on the hiring needs of the CFD, it was not likely that they would be called for further processing. However, the "qualified" pool was also told that "because it is not possible at this time to predict how many applicants will be hired in the next few years, your name will be kept on the eligible list maintained by the Department of Personnel for as long as that list is used."

On the same day that the City mailed the notice of scores to applicants, the City issued a press release detailing the results of the exam, including its disparate impact on minority applicants. Representatives from the League and a number of class plaintiffs met with counsel to discuss the legal implications of the 1995 Test. During the following year, plaintiffs' counsel obtained technical information from the City regarding the test's development and validation, which plaintiffs' experts reviewed. Based on the results of this analysis, several plaintiffs filed charges of discrimination with the EEOC. Plaintiffs then filed this lawsuit in September of 1998, seeking damages under Title VII for the City's unlawful use of the 1995 Test in its firefighter hiring practices.

ANALYSIS

I. Pre-Trial Motions

Before turning to the merits of plaintiffs' Title VII claim, the court will briefly address two preliminary matters.

First, the City has challenged the standing of the League as a plaintiff, arguing that the League: (a) is not a proper plaintiff under Title VII; and (b) does not otherwise meet the constitutional requirements for standing: injury in fact, causation, or redressability. *See Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560-61 (1992). The court disagrees. It is undisputed that the League is a non-profit organization made up of African-American firefighters, which, among other activities, seeks to recruit additional African-Americans to the CFD, increase African-American representation in the CFD, train African-American members of the CFD for promotional exams, and fight racism within the CFD. The disparate impact of the 1995 Test on African-American firefighter candidates has caused the League to suffer a concrete injury: decreased membership as a result of fewer African-Americans being hired for the position of firefighter. Additionally, the remedies available to the class plaintiffs under Title VII, particularly a hiring remedy, will likely redress the League's injury because more African-American firefighters means more potential members for the League. Moreover, because one of the primary aims of the League is to combat discrimination against African-Americans in the CFD, prevailing in this action will further the mission of the League. The type of injury suffered by the League and its likelihood of redress if it prevails is sufficient to justify the League's standing as a plaintiff in this case.

Second, plaintiffs have renewed their motion for judicial estoppel which was denied by the court without prejudice prior to trial.[4] Plaintiffs argue that the City should be judicially estopped from seeking to establish facts regarding the 1995 Test which are contrary to factual positions upon

[4] At the time the court denied plaintiffs' motion, it did not have the benefit of hearing the parties' theories of the case or their evidence in support, and did not believe it was in a position to rule on plaintiffs' motion. Now, of course, the court is well aware of the City's defense to plaintiffs' claims and can properly evaluate whether the City should be estopped from seeking to establish facts that appear contrary to those relied on in *Horan*.

which the City prevailed in another case involving that test, *Horan v. City of Chicago*, No. 98 C 2850, 2003 U.S. Dist. LEXIS 17173 (N.D. Ill. Sept. 30, 2003).

In *Horan*, white incumbent firefighters challenged a series of CFD affirmative action personnel decisions made by the City. As here, the parties' positions in *Horan* focused, in part, on their characterization of the results of the 1995 Test. In challenging the affirmative action decisions of the City, the *Horan* plaintiffs attempted to prove "that the 1995 entrance examination was content valid" and that firefighters with scores of 89 and higher on the 1995 entrance examination were better qualified than those with lower passing scores. *Horan*, 2003 U.S. Dist. LEXIS 17173, at *185. During the bench trial and in its proposed findings of fact submitted after trial, the City contested that argument and took positions that appear to question the validity of the 1995 Test and, therefore, ostensibly undermine positions taken by the City in this case.

The City's defense in *Horan* was that the 1995 Test could not predict overall firefighter performance. The City argued that the job of firefighter depended on proficiency in a number of physical, psychological, emotional, and cognitive skills and abilities, and that the 1995 Test, itself a measure of only a narrow set of cognitive abilities, could not predict on-the-job performance. In so arguing, the City took factual positions that, at least in some respects, are inconsistent with positions it has advanced in this case. For example, whereas the City now claims that the 1995 Test is a valid predictor of at least some aspects of firefighter performance or trainability, the City in *Horan* asserted that there was no evidence that those applicants who scored 89 and above on the 1995 Test were any better qualified to perform the job of firefighter than individuals who obtained a score between 65 and 88. Moreover, whereas the City now claims that success on the 1995 Test is an indicator of overall cognitive ability, the City in *Horan* argued that there are numerous

cognitive abilities required by the firefighter position that are not measured by the 1995 Test.

Plaintiffs in this case argue that the City ought to be estopped from switching tack from their prevailing position in *Horan*. "When a party assumes a certain position in a legal proceeding, and succeeds in maintaining that position, he may not thereafter, simply because his interests have changed, assume a contrary position." *New Hampshire v. Maine*, 532 U.S. 742, 749-51 (2001). "The purpose of the doctrine ... is to reduce fraud in the legal process by forcing a modicum of consistency on a repeating litigant." *Ladd v. ITT Corp.,* 148 F.3d 753, 756 (7th Cir. 1998). In other words, "a party who prevails on one ground in a lawsuit cannot turn around and in another lawsuit repudiate the ground. If repudiation were permitted, the incentive to commit perjury and engage in other litigation fraud would be greater. A party envisaging a succession of suits in which a change in position would be advantageous would have an incentive to falsify the evidence in one of the cases, since it would be difficult otherwise to maintain inconsistent positions." *McNamara v. City of Chicago,* 138 F.3d 1219, 1225 (7th Cir. 1998) (citations omitted).

Although it is a close question, after hearing the City's evidence in this case and comparing it to the City's prevailing positions in *Horan*, this court concludes that judicial estoppel is not applicable. Here, the City does not argue (or at least has not attempted to prove) that the 1995 Test accurately predicts overall job performance. Rather, the City's position appears to be that the 1995 Test predicts performance on a few of the cognitive aspects of the job related to "trainability." As discussed below, that position is not adequately supported and, in any event, is contrary to the City's obligations under Title VII. However, the court will not go so far as to hold the City estopped from espousing this argument. While the City's position in *Horan* may severely undermine its defenses in the instant case, its position is sufficiently different from its position in *Horan* to avoid estoppel.

That said, the court agrees with plaintiffs that factual assertions made by the City to the court in *Horan*, to the extent they are relevant in this case, are admissible as party admissions under Fed. R. Evid. 801(d)(2). Thus, the court admits into evidence Plaintiffs' Exhibit 61, which contains numerous proposed findings of fact submitted by the City after its trial in *Horan*. As discussed below, the admissions in *Horan* expose the weaknesses in the City's defenses in this case.[5]

II. The Merits Of Plaintiffs' Title VII Claim

The court now turns to the merits of plaintiffs' discrimination claim. Title VII employs a burden-shifting approach for disparate impact claims, which requires plaintiffs to prove first that the challenged, facially-neutral employment practice had a disparate impact on a protected class of people. 42 U.S.C. § 2000e-2(k)(1)(A)(i). In this case, the disparate impact of the 1995 Test is not in dispute; the parties have stipulated that the 1995 Test, used with a cut-off score set at 89, had a severe disparate impact on African-American firefighter candidates. Therefore, the burden of proof in this case shifts to the City to prove that its use of the 1995 Test was "job related for the position in question" and "consistent with business necessity." 42 U.S.C. § 2000e-2(k)(1)(A)(i). If the City justifies the adverse impact of the 1995 Test, the burden shifts back to plaintiffs to prove that a substantially equally valid, and less discriminatory alternative to the challenged practice was available but not employed. 42 U.S.C. § 2000e-2(k)(1)(A)(ii).

The 1991 Civil Rights Act defines the City's burden of proof, codifying the concepts of job relatedness and business necessity "enunciated by the Supreme Court in *Griggs v. Duke Power Co.*,

---

[5] In addition to its other pre-trial motions, the City has filed a motion to introduce supplemental authority related to the timeliness of plaintiffs' claims. The court has already granted that motion. However, to the extent the City's additional motion also seeks reconsideration of the court's order denying the City summary judgment on this issue, the motion is denied.

401 U.S. 424 (1971), and in other Supreme Court decisions prior to *Wards Cove Packing Co. v. Antonio*, 490 U.S. 642 (1989)." Pub. L. 102-166 § 3. The Seventh Circuit has clarified this standard, holding that "*Griggs* does not distinguish business necessity and job relatedness as two separate standards. It states that: 'The touchstone is business necessity. If an employment practice which operates to exclude [a protected group] cannot be shown to be related to job performance, the practice is prohibited.'" *Bew v. City of Chicago*, 252 F.3d 891, 894 (7th Cir. 2001) (quoting *Griggs*, 401 U.S. at 431). In other words, an employment test shown to have a disparate impact is presumptively unlawful unless it "bear[s] a demonstrable relationship to successful performance of the jobs for which it was used." *Griggs*, 401 U.S. at 431.

To prevail in this case, therefore, the City must prove that its decision to hire only those applicants who scored 89 and above on the 1995 Test was "predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated." *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 431 (1975). The critical question here is not so much whether the 1995 Test actually measures skills that are part of the job of firefighter, but whether setting the cut-off score at 89 "properly discriminate[d] between those who can and cannot perform the job well." *Bew*, 252 F.3d at 895; *Allen v. City of Chicago*, No. 98 C 7673, 2002 U.S. Dist. LEXIS 18973, at *10 (N.D. Ill. Sept. 30, 2002) (explaining that "[t]ests are valid if, and only if, they predict performance").

The court finds that, by that standard, the City has failed to prove that its use of the 1995 Test with a cut-off score of 89 was justified by business necessity and, therefore, the City's Title VII defense cannot succeed. The City's "business necessity" defense hinges on two central arguments: (1) that the 1995 Test is an effective measure of the applicants' relative abilities as to four specific

cognitive skills; and (2) an applicant's performance on the 1995 Test, at least in some respects, can predict (or correlates to) that applicant's performance on certain aspects of the job of Chicago firefighter. As explained below, the City's proof falls short on both arguments. The evidence at trial demonstrated that: (a) there are serious questions regarding whether the 1995 Test can reliably measure the four cognitive skills it was designed to measure; (b) the cut-off score of 89 is statistically meaningless in that it fails to distinguish between candidates based on their relative abilities; and (c) even assuming that the 1995 Test reliably measures the skills it is supposed to measure (and that the 89 cut-off score is a meaningful benchmark), the City failed to prove that test results could be used to predict firefighter performance, *i.e.,* that those who scored 89 or higher on the 1995 Test were more qualified for the job than those who scored between 65 and 89. In short, the court finds that the City has failed to prove that its selection process – which disproportionately excluded African-American applicants from the firefighter candidate pool – was justified by business necessity. Therefore, the court holds that selection procedure unlawful under Title VII.

A. The Ability Of The 1995 Test To Reliably Measure The Cognitive Skills That It Was Designed To Measure.

Before reaching the question whether the 1995 Test can accurately distinguish between those who can perform the job of firefighter and those who cannot, the court must address the threshold question whether the City has proven that the 1995 Test can reliably measure the four cognitive skills that it was designed to measure. After all, if the 1995 Test cannot even measure the cognitive skills in question, the City cannot reasonably claim that its reliance on the 1995 Test was justified by business necessity. The court has serious concerns regarding the City's proof on this threshold question.

The statistical reliability of the 1995 Test was established at trial; Dr. Outtz testified that 23% of the variance in an individual's score could be blamed on random error. Although that figure indicates that the 1995 Test is a relatively blunt instrument, the 1995 Test's reliability coefficient is within the acceptable range. However, the court's concerns are more fundamental: regardless of the effect of random error, it is not clear that the 1995 Test measures what it is supposed to measure. Rather, the evidence at trial indicated that design flaws in the video portion of the 1995 Test may have significantly affected the 1995 Test's ability to measure some of the cognitive skills at issue.

The video demonstration section was an entirely novel test, created in the hope that using an audiovisual component in the 1995 Test might minimize adverse impact. Like the rest of the 1995 Test, the video portion had never been "piloted" in a practical setting before its debut, was never used prior to the 1995 Test and has not been used since. According to Dr. Outtz, the video section – which represented 85% of the applicant's total score – was designed to measure three cognitive skills: an applicant's ability to: (a) understand oral instructions; (b) take notes; and (c) learn from or understand based on demonstration. Those skills were not measured by any other portion of the 1995 Test. However, the evidence at trial demonstrated that, contrary to that design, the results of the video portion of the 1995 Test hinged almost entirely on a single skill – the candidate's ability to take notes. Information in the video portion of the 1995 Test is complex, involves fictitious subject matter and is presented very quickly. Based on testimony from plaintiffs' expert, Dr. Cranny, as well as the court's own observation of the video demonstration, the court is persuaded that, aside from those test-takers blessed with a photographic memory, performance on the video portion of the 1995 Test depends on the applicant's ability to take effective notes while not missing any of the information conveyed by the video. The video demonstration is chaotic and is 83 minutes long. The

questions asked of candidates at the end of the demonstration require the candidates to recall specific facts from the 83 minute demonstration. If a candidate does not take voluminous and accurate notes during those 83 minutes, that candidate will perform poorly on that section regardless of his or her other cognitive abilities.

That design flaw is compounded by the fact that, even according to the City's own job analysis, the ability to take notes is not particularly important in performing the job of firefighter. The job analysis performed for the 1995 Test revealed that "note-taking" was dead last among the 46 identified abilities required for the job of Chicago firefighter. In fact, two subsequent job analyses for the position of San Francisco firefighter, performed in 1996 and 2000, failed to identify "note-taking" as a skill required by the position at all.

In short, the evidence at trial reflected that, contrary to the intentions of the 1995 Test's designers, the 1995 Test was skewed towards one of the least important aspects of the firefighter position at the expense of more important abilities. That fact undermines the 1995 Test's utility as a valid measure of candidates' relative cognitive skills and, therefore, undermines the City's defense in this case.

B. Inability Of The 89 Cut-Off Score To Distinguish Between Qualified And Unqualified Candidates.

As stated above, the keystone of the City's "business necessity" defense in this context is whether the City's selection strategy could distinguish between those qualified to be a firefighter and those who are not qualified for that position. However, the uncontradicted evidence at trial established that, contrary to that standard, the City's cut-off score of 89 could not – and was never intended to – make that distinction.

To survive a disparate impact challenge, "[A] discriminatory cutoff score on an entry level employment examination must be shown to measure the minimum qualifications necessary for successful performance of the job in question." *Lanning v. Southeastern Pennsylvania Transp. Authority (SEPTA)*, 181 F.3d 478, 481 (3d Cir. 1999); *United States v. Delaware*, No. Civ. A. 01-020-KAJ, 2004 WL 609331, at *24 (D. Del. Mar. 22, 2004) (explaining that "minimum qualifications necessary" means "likely to be able to do the job"). As interpreted by the Seventh Circuit, this means that a cut score may satisfy the business necessity requirement if it is based on "a professional estimate of the requisite ability levels, or, at the very least by analyzing the test results to locate a logical break-point in the distribution of scores." *Gillespie v. Wisconsin*, 771 F.2d 1035, 1045 (7th Cir. 1985).[6] The cut-off score of 89 in this case simply does not satisfy those criteria.

In fact, the cut score of 89 was a statistically meaningless benchmark; it provided no information regarding the relative abilities of the test-takers. As explained above, *supra*, pp. 8-9, because of the 1995 Test's large margin of error, Dr. Outtz – the 1995 Test's creator and one of the City's expert witnesses in this case – proposed scoring the 1995 Test using a sliding band of 13 points from the highest score of 98. Dr. Outtz made that proposal because he could not find any statistical difference between scores that are within 13 points of each other. Dr. Outtz testified that, because of the significant rate of error inherent in the 1995 Test, a cut-off score of 89 had no psychometric basis, meaning, there was no basis for an inference that people who had a higher score

[6] The EEOC's Uniform Guidelines – which are "entitled to great deference" by the court, *Albemarle*, 422, U.S. at 431 – provide that "where cut-off scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force." 29 C.F.R. § 1607.5(H); *Bew*, 252 F.3d at 894 (using "the EEOC's standard" to determine propriety of cut score in Title VII case).

within the 13-point band possessed more of the abilities measured by the 1995 Test than people who scored at the lower end of that range. Dr. Outtz informed the City of the shortcomings of the 1995 Test, notifying the City of the 1995 Test's 13-point margin of error and warning that there was no statistical basis for setting the cut-off score within that 13-point band.

The evidence demonstrated that the City ignored Dr. Outtz's counsel and set the cut score at 89 simply to limit the number of candidates selected for further processing. As the City admitted in *Horan*, the "cut score was not set by the City because it believed that individuals who scored 89 or higher were the best qualified candidates for the job of firefighter." Rather, the cut-score was established for "administrative convenience."

Based on Dr. Outtz's uncontroverted testimony about the statistical properties of the 1995 Test, the court finds that the City has not presented sufficient evidence to justify its admittedly discriminatory decision to set the cut score for the 1995 Test at 89 points. The evidence in this case clearly showed that the City: (1) knew that a cut-off score of 89 would disproportionately exclude African-American applicants from the candidate pool; and (2) knew that the cut-off score was a statistically useless method of evaluating candidates. However, ignoring the statistical limitations of the 1995 Test, the City went ahead and applied the 89 cut-off score for reasons of "administrative convenience" even though less discriminatory, and equally convenient, selection methods were available and later employed (namely, selection of applicants at random from the pool of candidates who passed the 1995 Test). Those facts alone are fatal to the City's defense in this case and lead the court to find that defendant's selection methods are unlawful under Title VII.

C. Validity / Predictive Value Of The 1995 Test.

Even assuming that the 1995 Test reliably measured the four cognitive abilities that it was

designed to measure (and ignoring the fact that the 89 cut-off score is statistically meaningless), the City's "business necessity" defense must fail because the City failed to prove that the 1995 Test, applied with a cut-off score of 89, can meaningfully distinguish candidates who are qualified to perform the job of firefighter from those who are not qualified for that position. As the City conceded in *Horan*, there "is no evidence to support a finding that the top seven (7) percent of the candidates on the written portion of the 1995 entrance examination [*i.e.*, those who scored 89 or above on the 1995 Test] are the most qualified candidates for the job or that they are better qualified than individuals who obtained a score between 88 and 65 . . ." That admission accurately summarizes the fatal weakness of the City's position in this case.

The ability of the 1995 Test to predict firefighter performance is key to the City's Title VII defense. "The mere fact that a test 'is representative of important aspects of performance on the job' (as content validity requires) matters only because it is reasonable to suppose that such a test will usefully distinguish among candidates – in other words, that using the test in selection will likely lead to a better performing workforce." *Allen*, 2002 U.S. Dist. LEXIS 18973, at *10.

The evidence in this case does not support such a supposition. As the City admitted in *Horan*, there is no evidence that candidates with a score of 89 and above are more qualified than those who passed the exam but fell short of the 89 cut score. The City has hired hundreds of paramedics and veterans who scored below an 89 on the 1995 Test. Moreover, most of the cadets who graduated from the Academy in 2003 scored between a 65 and 88. The City has presented no evidence that those firefighters are any less qualified on any aspect of job performance than those who scored 89 or above on the exam. To the contrary, the City has admitted a lack of correlation between test scores and job performance in the context of the 1995 Test's disparate impact on

African-Americans; the City admitted in *Horan* that both the designer of the 1995 Test, Dr. Outtz, and several of the CFD's top officials concluded that "there are no measured differences in job performance between Blacks and whites in any rank in fire services despite measured differences on cognitive ability tests."

Plaintiffs' expert, Dr. Charles Cranny, convincingly articulated the City's problem in statistical parlance, explaining that the predictive value of the 1995 Test cannot be determined because there is no "correlated known value." Although the test scores are known, there is no actual evidence of a correlation between those test scores and job performance. According to Dr. Cranny, while the two variables could be plotted on a "scatter graph" and a regression line could be drawn to reflect a linear relationship between test scores and job performance, without evidence of a correlation between the two variables (called the correlation coefficient), the strength of the relationship between test scores and job performance cannot be determined.

The City has attempted to overcome the dearth of evidence in this case, and its fatal admissions in *Horan*, by arguing that: (1) even if the 1995 Test is not predictive of overall job performance, it is a valid measure of the "trainability" of cadets; and (2) while there is no data in this case linking test performance and job performance, the 1995 Test should nevertheless be found valid because there is always a strong correlation between the results of cognitive tests and subsequent job performance. The court is not convinced by either of the City's arguments.

In support of its argument that the 1995 Test is a valid measure of the "trainability" of candidate firefighters, the City offered the testimony of Chicago Fire Chief and Assistant Director of Training, Steve Chikerotis. Chief Chikerotis testified that, in his opinion, the 2002 class of cadets who entered the Academy with scores between 65 and 88 performed less well on weekly pencil and

paper quizzes and needed more remedial work than prior cadet classes who entered the Academy with scores of 89 and above. The Chief's opinion on the relative "trainability" of cadets is based on his own observations of cadets training at the Academy.

The court finds that the City's evidence is insufficient to establish a relationship between test performance and the "trainability" of cadets. At the outset, the 1995 Test was not designed to measure skills related to trainability. In identifying the skills required of a Chicago firefighter, the designers of the 1995 Test focused exclusively on on-the-job observations. They did not conduct any observations of skills needed during training at the Academy. Moreover, while the court credits the testimony of Chief Chikerotis, the court finds that it is entitled to only modest weight. The court is uncomfortable relying on anecdotal evidence of training performance to prove an essential element of the City's defense, especially when the observations at the core of that anecdotal testimony occurred in the late stages of this litigation. In addition, Chief Chikerotis's testimony regarding cadets' performance on written exams did not provide a comprehensive picture of the cadets' training regimen. Chief Chikerotis made it clear that cadets are evaluated at the Academy on much more than their performance on quizzes and tests. Among other criteria, cadets are evaluated on their ability to operate fire engines, to perform rescues from multiple story buildings and to work as a team. The Chief testified that these skills and many others are essential to the job of firefighter and that candidates who fail to master those skills, regardless of their ability on written tests, will not pass the Academy. Since the cadets' scores on written tests do not reflect how well the cadets mastered the myriad other skills required to pass the Academy, those scores alone are not convincing

proof of the candidates' relative "trainability."[7]

In addition to its use of anecdotal evidence of the relative trainability of firefighter candidates, the City argues that the 1995 Test is valid for the simple reason that cognitive tests, as a general rule, are predictive of job performance. In essence, the City argues that the problem identified by Dr. Cranny (the lack of any correlation coefficient specific to the 1995 Test) can be overcome by borrowing correlation coefficients measured in other cognitive exams. The City's expert, Dr. Campion, testified that, although there is no data that links performance on the 1995 Test to job performance or "trainability," the City can rely on the correlation coefficients measured in other cognitive tests and use them to validate the 1995 Test regardless of whether those other tests measured any of the four cognitive skills that the 1995 Test was designed to measure. Dr. Campion's opinion is based on his review of 13 meta-analyses of general intelligence tests.[8] His resulting conclusion is that "cognitive abilities tend to correlate" in that "you can have widely different kinds of abilities, but yet they will correlate amongst each other in a reasonably representative sample of people." In other words, all cognitive tests are created equal and any well-designed cognitive test can be used to predict job performance.

While the court appreciates the value of meta-analysis to the field of industrial and organizational psychology in general, the court is not persuaded by the City's sweeping application of meta-analysis in this instance. Significantly, the City's broad conclusion that "all cognitive

---

[7] Chief Chikerotis also testified that the Academy switched its curriculum in the Fall of 2002. That switch in curriculum may well have accounted, albeit to some entirely unknowable degree, for some of the variance in performance to which the Chief testified.

[8] A meta-analysis is a statistical analysis of the results of a collection of individual studies to integrate and summarize their results.

abilities correlate" is strikingly different from its admissions in the *Horan* case, where the City emphasized that cognitive skills are varied and distinguishable and that the results – and consequently the predictive value – of a cognitive test can vary depending on which skills are tested.[9] The testimony of plaintiffs' expert, Dr. Cranny, is consistent with the City's position in *Horan*, and the court finds the City's position in *Horan,* and not its argument here, more persuasive. Even accepting that there is *some* correlation between various tests of cognitive and mental abilities, that hardly establishes that those tests test substantially the same thing or are interchangeable.

The 1995 Test was unique. It was designed to measure only four specific cognitive abilities and included a heavily-weighted video demonstration section that was never piloted and was never used before or since. As discussed above, the unique structure of the 1995 Test was far from perfect and may have interfered with the 1995 Test's ability to measure some of the skills it was intended to measure. Yet, regardless of the 1995 Test's unique design and evident flaws, the City would have the court import data from other cognitive tests based on the simple conclusion that "all cognitive abilities correlate." The City asks the Court to reach this conclusion without evidence or analysis of whether the tests underlying the City's conclusions are comparable to the 1995 Test. Given the unique character of the test at issue here, and the lack of evidence of the nature of the tests on which the meta-analytic studies discussed by Dr. Campion were based, the court rejects the City's argument that those studies validate the 1995 Test.

D. Less Discriminatory Alternative.

The Court finds that the City did not carry its burden of proof in this case and, therefore, rules

[9] Defendant's other expert, Dr. Outtz, was far more circumspect on this point than Dr. Campion, noting that cognitive abilities correlate only "for the most part" and sometimes do not.

in favor of plaintiffs on their Title VII claim. However, even if the City had successfully proven that the disparate impact of its decisions was justified by business necessity – and thereby shifted the burden of proof back to plaintiffs – plaintiffs would still prevail in this case because the evidence clearly shows that an equally valid and less discriminatory alternative was available. *See* 42 U.S.C. § 2000e-2(k)(1)(A)(ii) (describing burden shifting standard).

Quite simply, the City could have done what it is doing now: it could have randomly selected candidates who passed the exam for further evaluation. Such an alternative would have been less discriminatory; although the 1995 Test would have had a disparate impact on African-American candidates regardless of the cut-off score, random selection of qualified candidates has indisputably lessened the disparate impact of the 1995 Test. Moreover, the new policy of random selection of qualified applicants is "equally valid" in that it is equally effective at serving the essential goal of the CFD, producing quality firefighters.[10] There is no indication that the shift in selection procedures caused a drop-off in the quality of firefighters produced by the Academy. As the City candidly admitted in *Horan*, there is no evidence that firefighters who scored between 65 and 89 are any less qualified than candidates who scored 89 or above.

The court finds that, from 1995 to 2001, the City used a hiring procedure that had a disparate impact on African-American candidates even though an equally valid, and less discriminatory, option was available. For that reason – even if the City had proven that its practice was justified by business necessity – plaintiffs are entitled to a ruling in their favor on the liability aspects of their

[10] The new random selection policy also serves the City's stated goal of "administrative convenience." With random selection from the pool of qualified candidates, the City, without further deliberation or administrative action, can meet its hiring goals without clogging the process with an unmanageable number of candidates.

Title VII claims.

CONCLUSION

The City admits that its use of the 1995 firefighter examination with a cut score of 89 had a disparate impact on African-American applicants, and has failed to prove that its hiring procedures were job-related and consistent with business necessity. The court therefore concludes that the City's use of the 1995 Test with a cut-off score of 89 was a manifest violation of Title VII and enters judgment of liability against the City of Chicago and in favor of plaintiffs.

ENTER:

/s/
JOAN B. GOTTSCHALL
United States District Judge

DATED: March 22, 2005