ORIGINAL

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS
DALLAS DIVISION

UNITED STATES OF AMERICA,	§	
	§	
Plaintiff,	§	
	§	
v.	§	Civil Action No. 3:98-CV-0307-L
	§	
CITY OF GARLAND, TEXAS,	§	
	§	
Defendant.	§	

MEMORANDUM OPINION AND ORDER

Plaintiff United States of America ("Plaintiff" or "the United States") filed this action against Defendant City of Garland, Texas ("Defendant" or "the City") on February 6, 1998. This case involves a disparate impact claim under Title VII of the Civil Rights Act of 1964 ("Title VII"), 42 U.S.C. § 2000e *et seq.*, in which the United States challenges the City's employment practices in hiring entry-level police officers and firefighters. The court conducted a two-week bench trial from August 13, 2001 to August 24, 2001, on Plaintiff's claims. Plaintiff filed its Proposed Findings of Fact and Conclusions of Law on November 15, 2001, and Defendant filed its Proposed Findings of Fact and Conclusions of Law on November 20, 2001.[1] The court now rules in favor of the City

[1]Plaintiff filed its Objections to Defendant's Proposed Findings of Fact and Conclusions of Law on December 13, 2001, and Defendant filed its Objections to Plaintiff's Proposed Findings of Fact and Conclusions of Law on December 14, 2001. Plaintiff subsequently sent the court a letter dated March 22, 2004, to which it attached a copy of the "Post-Trial Findings of Fact and Conclusions of Law," issued in a case filed in the United States District Court for the District of Delaware, case number 01-020-KAJ, styled *United States v. The State of Delaware, The Delaware Department of Public Safety, and The Delaware Division of State Police*. Defendant responded in a letter received on March 26, 2004. The court hereby **orders** the clerk of the court to file Plaintiff's letter and attachment, dated March 22, 2004, and Defendant's letter, received March 26, 2004. The court reviewed the Delaware court's opinion and notes that the Delaware court concluded what this court also concludes, namely, that the ALERT examination is a valid and reliable test and that a cutoff score of up to 70% is reasonable.

on Plaintiff's claims. Based on a preponderance of the evidence and the applicable law, the court makes the following findings of fact and conclusions of law as required by Fed. R. Civ. P. 52(a).[2]

I. Findings of Fact[3]

A. Background

The City is a Home-Rule municipality under the Council-Manager form of government. The City Council is comprised of eight members who are elected by qualified voters within their respective districts, and the Mayor who is elected at large. Council members are elected to two-year terms and are limited to two consecutive terms.

The City, which was founded in 1891, is the ninth largest city in Texas and one of the 100 largest cities in the United States. It covers 57 square miles in northeast Dallas County including 243 acres of forests and 1,689 acres of parks. At the time of trial, the City had a population of 203,688 and consisted of more than 4,500 businesses.

The employment practices at issue in this case all relate to the use of three different cognitive skills tests used by the City to hire entry-level police officers and firefighters: the ALERT entry-level

---

[2]The court apologizes for the delay in issuing its findings of fact and conclusions of law. As the parties are aware, this litigation has been both protracted and difficult, requiring review of voluminous, highly technical documents and testimony.

[3]The facts contained herein are either undisputed or the court has made the finding based on the credibility or believability of each witness. In doing so, the court considered all the circumstances under which the witness testified which include: the relationship of the witness to Plaintiff or Defendant; the interest, if any, the witness has in the outcome of the case; the witness's appearance, demeanor, and manner of testifying while on the witness stand; the witness's apparent candor and fairness, or the lack thereof; the reasonableness or unreasonableness of the witness's testimony; the opportunity of the witness to observe or acquire knowledge concerning the facts to which he or she testified; the extent to which the witness was contradicted or supported by other credible evidence; and whether such contradiction related to an important factor in the case or some minor or unimportant detail. When necessary, the court comments on the credibility of a witness or the weight to be given a witness's testimony.

police officer examination;[4] the International Personnel Management Association ("IPMA") B-4 entry-level firefighter examination (used in April 1992); and the ALARM entry-level firefighter examination (used since 1994).[5] From 1992 through 1999, Neil R. Barnett, the City's Civil Service Director, decided which entrance examinations were administered to the police officer and firefighter applicants. The ALERT,[6] the IPMA B-4,[7] and the ALARM[8] examinations are paper-and-pencil multiple choice tests designed to measure certain cognitive skills. None of these examinations was

---

[4]Since 1992, the City used the following ALERT examinations to select entry-level police officers:

Test Developer	Test Series	Date
WOLLACK	ALERT 02/BB	July 18, 1992
WOLLACK	ALERT A-3	May 21, 1993
WOLLACK	ALERT A-4	July 8, 1994
WOLLACK	ALERT A-5	October 6, 1995
WOLLACK	ALERT A-6	August 16, 1996
WOLLACK	ALERT A-1	October 3, 1997
WOLLACK	ALERT A-2	February 20, 1998
WOLLACK	ALERT A-4	May 10, 1999
WOLLACK	ALERT A-3	May 14, 1999
WOLLACK	ALERT A-5	June 26, 1999
WOLLACK	ALERT A-4	May 19, 2000
WOLLACK	ALERT A-7	July 7, 2000

[5]Since 1992, the City used the following examinations to select entry-level firefighters:

Test Developer	Test Series	Date
IPMA	B-4	April 11, 1992
WOLLACK	ALARM 01	March 4, 1994
WOLLACK	ALARM 02	August 18, 1995
WOLLACK	ALARM 03	September 6, 1996
WOLLACK	ALARM 04	August 21, 1997
WOLLACK	ALARM 01	August 14 and 18, 1998
WOLLACK	ALARM 02	August 17 and 20, 1999
WOLLACK	ALARM 03	June 16, 2000

[6] The ALERT examination is a 160-item test with seven alternate forms.

[7]The IPMA B-4 examination is a 85-item test.

[8]The ALARM examination is a 100-item test with four alternate forms.

intended to assess noncognitive traits; instead, these traits are assessed in the post-test components of the selection process.

The City uses what is sometimes referred to as a "multiple hurdle" model, in which applicants are weeded out at several different stages in the selection process. The first hurdle is a preliminary screening for basic prerequisites (age, citizenship, education, and so forth).[9] The second hurdle is one of the cognitive skills tests listed above; applicants scoring below 70% on the test are rejected.[10]

---

[9]From 1992 to 1995, to meet the minimum qualifications for the position of police officer, an applicant had to:

(1) be 21 years of age on the test date, or at least 18 years of age and have 60 hours of college credit or 2 years of military experience;
(2) have a high school diploma or a GED;
(3) provide a social security number or driver's license number;
(4) be a United States citizen;
(5) return a completed application prior to the filing deadline;
(6) sign the application;
(7) provide a Form DD214 if the applicant is a veteran and wants to have veteran preference points added to his examination score.

The minimum qualifications changed slightly from 1996 to 1998. Specifically, a police officer applicant had to be 18 years of age or older and have 60 hours of college credit or an Associates Degree; 30 hours of college credit and 2 years of active duty military service; or 30 hours of college credit and 2 years experience as a commissioned peace officer (full or part-time). All other minimum qualifications stayed the same. Moreover, beginning in 1997, police officer applicants could not be over the age of 35, or 44 if other requirements were met, as of the date of the examination.

Similarly, from 1992 through the date of trial, an applicant for the position of firefighter had to:

(1) be 18 years old on the date of the test;
(2) have a high school diploma or a GED;
(3) provide a social security number or driver's license number;
(4) be a United States citizen;
(5) return a completed application prior to the filing deadline;
(6) sign the application;
(7) provide a Form DD214 if the applicant is a veteran and wants to have veteran preference points added to his examination score.

Beginning in 1997, firefighter applicants could not be over the age of 35 as of the date of the examination.

[10]If an applicant scored 70% or higher and provided a Form DD214, five veteran preference points (expressed as a percentage) were added to the applicant's score.

The third hurdle, used only for applicants for firefighter positions, is described as a cost-savings measure. In the third hurdle, the Garland Fire Department ("Fire Department") processes approximately 30 to 50 test-takers based on their scores on the cognitive skills test. Additional applicants are processed on an as-needed basis. Only a small number of openings are typically available in the Fire Department, so only a limited number of applicants from any given test date will eventually be hired. Subsequent stages in the screening process are costly, and the City limits the number of applicants to be processed through these stages.[11]

Those applicants who pass the second and (if applicable) third hurdles undergo a physical agility test, given on a pass-fail basis. Applicants who pass the physical agility test are subjected to further screening on a pass-fail basis, including an oral interview board, comprehensive background investigation, psychological screening (police only), medical assessment, polygraph examination, and interview. The applicants proceed through this process in descending rank order based on their cognitive skills test scores, including veteran preference points, if any.

The final hurdle is the hiring selection itself.[12] Applicants who pass all components of the screening process are hired in descending rank order[13] based their cognitive skills test scores,

---

[11]The Fire Department typically has more applicants who pass the second hurdle than available openings. For example, between 1992 and 1998, nearly 1,800 applicants passed the second hurdle, but only 81 were ultimately hired.

[12]There is no allegation of post-test adverse impact. In other words, there is no allegation that among those applicants who passed the cognitive skills test, there was a statistically significant difference in the hiring rates of black, Hispanic (collectively "minority"), or white applicants in either department.

[13]Between 1987 and 1990, there were nine firefighter applicants who were hired out of rank order. Their scores on the cognitive skills test were between 63.5% and 87%, and they ranked between 88 and 420 on the eligibility lists.

including veteran preference points, if any.[14] Thus, if the City has more openings than remaining applicants, all applicants are hired.[15] If the City has more applicants than openings, some applicants who scored better than the 70% cutoff score will be rejected.[16]

Those hired also undergo extensive state-mandated training at the respective training academies[17] and must pass state certification examinations[18] before beginning field training. For example, a police officer recruit must pass the Texas Commission on Law Enforcement Standards and Education ("TCLEOSE") basic peace officer certification test in order to enter the field training phase. Similarly, a firefighter recruit must pass the TCFP basic structural firefighter certification test prior to beginning work as a probationary firefighter. Although more individuals may "wash out" or be eliminated at these later points, the court need not consider these processes, as the United States challenges no practices that occur after the final selection, described above, takes place.

---

[14]Veteran preference points did not adversely affect minority applicants; instead, the points statistically worked to their advantage.

[15]The number of openings is fixed by budgetary appropriation based on need. Specifically, the City combines estimated vacancies due to retirements and other attrition with the historic dropout and failure rates at the police or fire academy and probationary training to arrive at the number of entry-level police officers and firefighters, respectively, it can hire in a given year.

[16]Since 1992, the Fire Department has never exhausted an eligibility list, as it has had enough candidates to fill the number of positions available. The Garland Police Department ("Police Department"), on the other hand, has exhausted all of its eligibility lists, except for those in 1993 and 1994, thus hiring all applicants who successfully completed all of the hurdles.

[17]Police officer recruits attend the Garland Police Academy. Police academy grades are used to determine seniority among police officers hired on the same day. Firefighter recruits, who are not already certified by the Texas Commission on Fire Protection ("TCFP"), attend the Dallas Fire Academy. As firefigher recruits are required to become Emergency Medical Technicians ("EMTs"), they, as well as current firefighters who are promoted to the rank of driver and are not certified paramedics, must also attend paramedic school. Prior to February 2001, Garland firefighters who went to paramedic school attended the paramedic training program at the University of Texas Southwestern Medical Center at Dallas.

[18]Scores on the state certification examinations are used to ensure that a recruit has achieved certification.

During November 1993, the City began consulting with Dr. Stephen Wollack[19] in anticipation of and in preparation for this litigation on issues relating to the validation of and adverse impact, if any, resulting from the ALERT and ALARM examinations. The parties stipulated to a database to be used for this litigation which included information regarding individuals who were affected by the challenged employment practices used by the City during the relevant time period.

B. Expert Witnesses

1. Plaintiff's Expert Witnesses

a. Bernard Siskin

Dr. Bernard Siskin ("Dr. Siskin") testified regarding his analysis of the employment practices challenged by the United States. Dr. Siskin holds a doctorate in statistics for the University of Pennsylvania and is a former chair of the Department of Statistics of Temple University. Since 1984, Dr. Siskin has been a statistical and economic consultant. He has testified as an expert witness over a hundred times.

b. Frank Landy

Dr. Frank Landy ("Dr. Landy") testified regarding the broad range of cognitive and noncognitive attributes that are necessary to perform the jobs of a police officer and firefighter. Dr. Landy holds a bachelor's degree in psychology from Villanova University and master's degree and doctorate in industrial psychology from Bowling Green State University. He has been an industrial organizational psychologist for approximately 30 years and has many years of experience with entry-level and promotional examinations for police officers and firefighters.

---

[19]Wollack & Associates developed the QB entry-level police examination and has made it available to the public continuously since 1987.

c. David Paul Jones

Dr. David Paul Jones ("Dr. Jones") testified regarding the development and validation of the Protective Services Profile ("PSP"), an entry-level police officer selection device. Dr. Jones was the project director for Aon Consulting, the firm that developed and validated the PSP. Dr. Jones holds a doctorate in industrial organizational psychology from Bowling Green State University. He has over 25 years of experience in designing and validating employee selection systems for public and private employers, including employers of law enforcement personnel.

d. Leaetta Hough

Dr. Leaetta Hough ("Dr. Hough") testified regarding the broad range of cognitive and noncognitive attributes that are necessary to perform the jobs of a police officer and firefighter. Dr. Hough holds a doctorate in industrial organizational psychology. Within her field, Dr. Hough specializes in the area of personality. Dr. Hough has experience in designing and validating employment selection procedures, including job analyses and test development for entry-level police officers and firefighters.

e. Christopher Hornick

Dr. Christopher Hornick ("Dr. Hornick") testified regarding development and validation of the CWH Selections Solutions Firefighter Examination ("CWH examination"), a selection device for entry-level firefighters. Dr. Hornick is the president of CWH Management Solutions, the firm that developed and validated the CWH examination. Dr. Hornick holds a bachelor's degree in psychology from the University of Utah, a master's degree in psychology from Texas Christian University, and a doctorate also from Texas Christian University. He has more than twenty years of experience analyzing the work of entry-level firefighters.

2. Defendant's Expert Witnesses

a. Judith Stoikov

Dr. Judith Stoikov ("Dr. Stoikov") is a statistician and labor economist. Dr. Stoikov holds a bachelor's degree in industrial and labor relations from Cornell University and a doctorate in economics from the London School of Economics and Political Science. During the past 20 years, Dr. Stoikov has been recognized by federal and state courts as an expert in labor economics and statistical analysis.

b. F. Mark Schemmer

Dr. F. Mark Schemmer ("Dr. Schemmer") testified regarding the development and validation of the IPMA B-4 examination. Dr. Schemmer holds a bachelor's degree in psychology with a minor in mathematics, a master's degree in cognitive psychology, and a doctorate in quantitative psychology, all from the University of Maryland. He has over 20 years of experience in personnel test development and validation.

c. David Morris

Dr. David Morris ("Dr. Morris") testified regarding the necessity of cognitive skills for the job of a police officer and his firm's role in the revision of the IPMA B-4 examination. Dr. Morris is employed with the Morris & McDaniel, a firm specializing in industrial organizational practice, and has been so employed since 1976. Dr. Morris holds a bachelor's degree in psychology from Milsaps College, a master's degree in psychology from Mississippi State University, and a doctorate in psychology from the University of Southern Mississippi. He has extensive experience in personnel test development and validation.

d. Kathleen K. Lundquist

Dr. Kathleen K. Lundquist ("Dr. Lundquist") testified regarding the validation of the employment selection devices at issue in this case. Dr. Lundquist holds a bachelor's degree in psychology and a master's degree and a doctorate in psychology specializing in psychometrics. Dr. Lundquist is president of Applied Psychology Techniques. She has approximately 20 years of experience in employment test development and validation.

e. Stephen Wollack

Dr. Stephen Wollack ("Dr. Wollack") testified regarding the development and validation of the employment selection devices at issue in this case. Dr. Wollack holds a bachelor and master's degree from Temple University and a doctorate in industrial organizational psychology from Bowling Green State University. Dr. Wollack is vice president of Wollack & Associates, which publishes police and firefighter examinations. He has extensive experience in employment test development and validation.

f. Gerald V. Barrett

Gerald V. Barrett ("Dr. Barrett") testified regarding the validation of one of the alternative selection devices proposed by the United States. Dr. Barrett holds a bachelor's degree in psychology from Whittenberg University, a master's degree in psychology and a doctorate in industrial organizational psychology from Case Western Reserve University, and a doctorate of jurisprudence from Akron School of Law. He has extensive experience in the development and validation of employment selection devices

3. Standard for Admitting Expert Testimony

Fed. R. Evid. P. 702 and 703, as interpreted in *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993), and *Kumho Tire Co., Ltd. v. Carmichael*, 526 U.S. 137 (1999) govern the admissibility of testimony of expert witnesses. Rule 702 provides:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

Fed. R. Evid. P. 702. Accordingly, before allowing expert testimony to be heard, a district court must be assured that the proffered witness is qualified to testify by virtue of his "knowledge, skill, experience, training or education." A district court should refuse to allow an expert witness to testify if it finds that the witness is not qualified to testify in a particular field or on a given subject. *Wilson v. Woods*, 163 F.3d 935 (5th Cir. 1999). The issue is whether a particular expert has "sufficient specialized knowledge to assist the jurors in deciding the particular issues . . . ." *Tanner v. Westbrook*, 174 F.3d 542, 548 (5th Cir. 1999) (quoting *Kumho*, 526 U.S. at 156).

In evaluating the admissibility of expert testimony, the key factors are reliability and relevance. *Daubert*, 509 U.S. at 589 (under Rule 702, expert testimony must be "not only relevant, but reliable"). The relevancy requirement ensures that the expert testimony will actually "assist the trier of fact to understand the evidence or to determine a fact in issue." *Id.* Fed. R. Evid. 401 defines relevant evidence as that which has "any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence." The reliability requirement ensures that the expert testimony is "supported

by appropriate validation" and "establishes a standard of evidentiary reliability." *Id.* at 590. This determination requires the court to assess "whether the reasoning or methodology underlying the testimony is scientifically valid and . . . whether that reasoning or methodology properly can be applied to the facts in issue." *Id.* at 593. To assist courts in this determination, the Supreme Court has suggested that trial courts examine a nonexclusive list of factors including whether a theory or technique has or can be tested, published, subjected to peer review, has or can be subjected to standards controlling its operation, the known or potential rate of error, and whether the theory or technique is generally accepted. *Id.* at 593-94. The Court has also instructed trial courts to "be mindful of other applicable rules," such as Rules 703, 706, and 403, when determining the admissibility of expert testimony. *Id.* at 595.

Kumho stressed that the *Daubert* factors may be relevant to the reliability of experience-based testimony. The overarching goal of *Daubert's* gate-keeping requirement, however, is to ensure the reliability and relevancy of expert testimony and to make certain that an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field. *Kumho*, 526 U.S. at 152. The party offering the expert testimony has the burden of establishing that it is admissible under Rule 702 and the *Daubert-Kuhmo* standard. *See Kumho*, 526 U.S. at 147; *see also Mathis v. Exxon Corp.*, 302 F.3d 448, 459-460 (5th Cir. 2002).

This case involves a substantial amount of highly technical documents and testimony, and with eleven experts is really a battle between the heavyweights on each side. The court determines that all of the experts presented by both parties have the requisite educational background and experience and sufficient knowledge as to the matters to which they rendered an opinion. The court

further determines that the testing performed or the methodology employed was sufficiently reliable and relevant to pass muster under Fed. R. Evid. P. 702. The issue, therefore, is which expert's testimony is most persuasive or convincing.

C. Other Witnesses

1. Plaintiff's Witnesses

a. Jose Hernandez Jr.

Jose Hernandez, Jr. ("Officer Hernandez") testified that he took the ALERT examination administered by the City in 1995. He scored 68% which was too low to be hired by the City. Officer Hernandez then took a civil service examination for a police officer position with the City of Plano, which he passed. He was subsequently hired and is now a Plano police officer.

b. Steven Oliphant

Steven Oliphant ("Officer Oliphant") testified that he took the ALERT examination administered by the City in 1997. He scored 60.63% which was too low to be hired by the City. Officer Oliphant then took a civil service examination for a police officer position with the City of Dallas, which he passed. He was subsequently hired and is now a Dallas police officer.

c. Cornelius Gibson

Cornelius Gibson ("Officer Gibson") testified that he took the ALERT examination administered by the City in 1995. He scored 60% which was too low to be hired by the City. At the time he took the examination, Officer Gibson was a certified peace officer in Texas and had completed police academy training at Cedar Valley Police Academy in Lancaster, Texas. Officer Gibson is currently a deputy with the Dallas Marshal's Office.

d. Enrique Jemmott

Enrique Jemmott ("Lieutenant Jemmott") testified that he took the ALERT examination administered by the City in 1996. He scored 69.4% which was too low to be hired by the City. At the time he took the examination, Lieutenant Jemmott had worked as a police officer for nine years in San Diego, California. Lieutenant Jemmott is currently a lieutenant with the Richland College Police Department.

2. Defendant's Witnesses

a. Neil Reichenberg

Neil Reichenberg ("Mr. Reichenberg") testified regarding the IPMA B-4 examination. Mr. Reichenberg is the executive director of the IPMA.

b. Neil Barnett

Neil Barnett ("Mr. Barnett") testified regarding the employment selection devices used by the City. Mr. Barnett was the personnel director and civil service director for the City during the time at issue in this case.

c. Daniel Grammar

Daniel Grammar ("Chief Grammar") testified regarding his experience as a Garland firefighter and as a supervisor of firefighters. He also testified regarding the cognitive demands of a Garland firefighter. Chief Grammar is chief of the Garland Fire Department.

d. Christopher Balanciere

Christopher Balanciere ("Firefighter Balanciere") testified regarding his experience as a Garland firefighter and the cognitive demands of a Garland firefighter. Firefighter Balanciere is a firefighter/EMT with the Garland Fire Department.

e. David Holcombe

David Holcombe ("Deputy Chief Holcombe") testified regarding his experience as a Garland firefighter and as a supervisor of firefighters. He also testified regarding the cognitive demands of a Garland firefighter. Deputy Chief Holcombe is a deputy chief with the Garland Fire Department.

f. William Crews

William Crews ("Firefighter Crews") testified regarding his experience as a Garland firefighter and the cognitive demands of a Garland firefighter, including the cognitive demands of being a paramedic. Firefighter Crews is a firefighter/EMT with the Garland Fire Department.

g. Jeffrey Tokar

Jeffrey Tokar ("Captain Tokar") testified regarding his experience as a Garland firefighter and as a supervisor. He also testified as to the costs associated with the Fire Department's selection process. Captain Tokar is a captain with the Garland Fire Department.

h. Larry Wilson

Larry Wilson ("Chief Wilson") testified regarding testified regarding his experience as a police officer and as a supervisor of police officers. He also testified regarding the cognitive demands of the job of a Garland police officer. Chief Wilson is the chief of the Garland Police Department.

i. Brandon Hernandez

Brandon Hernandez ("Officer Hernandez") testified regarding his experience as a Garland police officer and the cognitive demands on Garland police officers. Officer Hernandez is a Garland police officer.

j. Lanny Orman

Lanny Orman ("Officer Orman") testified regarding his experience with the Garland Police Academy and the Police Department's selection process. Officer Orman is a Garland police officer.

k. Alexis Crockett

Alexis Crockett ("Officer Crockett") testified regarding his experience as a Garland police officer and specifically the field training officer program. Officer Crockett is a field training officer with the Garland Police Department.

l. Jack James

Jack James ("Captain James") testified regarding his experience with the Fire Department's selection process. Captain James is a captain with the Garland Fire Department and has been employed with the Fire Department since 1975.

m. William David Thurman

William David Thurman ("Officer Thurman") testified regarding his role in the Police Department's selection process. Specifically, Officer Thurman has been a background investigator for the Garland Police Department since 1995.

n. Terry Wayne Moore

Terry Wayne Moore ("Captain Moore") testified regarding the selection and hiring process used by the Missouri State Highway Patrol, including the use of the ALERT examination and an alternate compensatory examination purportedly recommended by the Department of Justice. Captain Moore is a captain with the Missouri State Highway Patrol. He has been employed with the Missouri State Highway Patrol since approximately 1977.

II. Conclusions of Law[20]

A. Disparate Impact

A disparate impact claim arises when an employer's otherwise neutral policy has a "significantly disproportionate impact" on a protected class. § 2000e-2(k)(1)(A)(i); *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975). Disparate impact claims do not require proof of intent to discriminate. *Munoz v. Orr*, 200 F.3d 291, 299 (5th Cir.), *cert. denied*, 531 U.S. 812 (2000). Instead, they focus on facially neutral employment practices that create such statistical disparities disadvantaging members of a protected group that they are "functionally equivalent to intentional discrimination." *Id.* (quoting *Watson v. Fort Worth Bank and Trust*, 487 U.S. 977, 987 (1988)). Disparate impact claims may be brought by individual plaintiffs or as a class action. *Munoz*, 200 F.3d at 299. In either case, however, the evidence must focus on the degree of statistical disparity between protected and non-protected workers, and the burden is on the plaintiff to conduct a systemic analysis of those employment practices and present evidence in order to establish his or her case.

The framework for proof of disparate impact claims involves a three-step process, with shifting burdens of proof:

> To establish a prima facie case of discrimination, a plaintiff must show that the facially neutral employment practice had a significantly discriminatory impact. If that showing is made, the employer must then demonstrate that "any given requirement [has] a manifest relationship to the employment in question," in order to avoid a finding of discrimination. Even in such a case, however, the plaintiff may prevail, if he shows that the employer was using the practice as a mere pretext for discrimination.

---

[20]Any finding of fact which may also constitute a conclusion of law shall be deemed a conclusion of law, and any conclusion of law which may also constitute a finding of fact.

Connecticut v. Teal, 457 U.S. 440, 446-47 (1982) (citations omitted). A showing that another employment practice would serve the employer's legitimate interests without a similar undesirable racial effect is evidence of pretext. *Albemarle*, 422 U.S. at 425.

1. Prima Facie Case

"A typical target of disparate impact analysis is an aptitude test used to screen applicants for promotion or hire." *Bunch v. Bullard*, 795 F.2d 384, 392 (5th Cir. 1986). A plaintiff can establish a prima facie case of disparate impact by showing "that the tests in question select applicants for hire . . . in a racial pattern significantly different from that of the pool of applicants." *Albemarle*, 422 U.S. at 425. Disparate impact claims necessarily rely heavily on statistical evidence. *Munoz*, 200 F.3d at 300. Statistical discrepancies in pass rates must be significant to establish a prima facie case. *Frazier v. Garrison Indep. Sch. Dist.*, 980 F.2d 1514, 1524 (5th Cir. 1993); *Hill v. Mississippi State Employment Serv.*, 918 F.2d 1233, 1238 (5th Cir. 1990), *cert. denied*, 502 U.S. 864 (1991).[21] "A plaintiff may use statistical as well as non-statistical evidence in establishing a prima facie case of disparate impact." *Page v. U.S. Industries, Inc.*, 726 F.2d 1038, 1053 (5th Cir. 1984). "While often helpful, standard deviation techniques are not the sole measure of the impact of a facially neutral requirement. We cannot permit an employer to escape liability for discriminatory tactics merely because his work force is not vast enough to provide meaningful data for a sophisticated statistical evaluation." *Bunch*, 795 F.2d at 395. Thus, a prima facie case does not require a sample size large enough for normal tests of statistical significance.

---

[21] *See also* 29 C.F.R. § 1607.4(D) ("A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact . . . ."). This is, of course, not the exclusive measure of whether a difference is significant.

In this case, in order to establish a prima facie case of disparate impact, the United States must (1) identify the challenged employment practice or policy; (2) demonstrate a disparate impact on a group that falls within the protective ambit of Title VII; and (3) demonstrate a causal relationship between the identified practice and the disparate impact. *Gonzales v. City of New Braunfels*, 176 F.3d 834, 839 n.26 (5th Cir.1999).

2. Job Relatedness and Business Necessity

If the United States meets its initial burden at the first step of the disparate impact analysis, the court then considers "the employer's justification for the use of these practices." *Frazier*, 980 F.2d at 1525-26. In other words, the burden shifts to the City to prove that the employment practice is valid by showing that it is "job related" and "consistent with business necessity." 42 U.S.C. § 2000e-2(k)(1)(A)(i). An employer can establish business necessity by showing that the selection criteria bear "a manifest relationship to the employment in question. *Griggs*, 401 U.S. at 432. The employer must also establish that its legitimate employment needs are significantly served by the challenged employment practice; however, it need not show that those needs require the employment practice. *New York City Transit Authority v. Beazer*, 440 U.S. 568, 587 n.31 (1979).

There are three methods of validating an employment practice: criterion-related validation, content validation, and construct validation. *Washington v. Davis*, 426 U.S. 229, 247 n.13 (1976); 29 C.F.R. § 1607.5(B). Criterion-related validation determines validity by "identifying criteria that indicate successful job performance and then correlating test scores and the criteria so identified." *Washington*, 426 U.S. at 247 n.13. In other words, criterion-related validation relies on a statistical correlation between good performance on the job, based on established criteria, and good

performance on the test, for example, whether the LSAT predicts good law school grades.[22] *See generally* 29 C.F.R. §§ 1607.5, 1607.14. Construct validity determines validity by structuring examinations "to measure the degree to which job applicants have identifiable characteristics that have been determined to be important in successful job performance." *Washington*, 426 U.S. at 247 n.13. In a construct validation, the test measures some abstract ability/quality, which is considered important to doing the job well; an IQ test is a common example. *See generally* 29 C.F.R. §§ 1607.5, 1607.14. Content validity determines validity with "tests whose content closely approximates tasks to be performed on the job by the applicant." *Washington*, 426 U.S. at 247 n.13. In other words, in a content validation, the test is essentially equivalent to some representative portion of the job, for example, a typing test for a secretary. *See generally* 29 C.F.R. §§ 1607.5, 1607.14.

3. Alternatives

If the City shows that the examinations and their uses are job-related and consistent with business necessity, the United States can still prevail by demonstrating "the availability of alternative practices to achieve the *same* business ends, with less racial impact." *Frazier*, 980 F.2d at 1525-26 (emphasis added); 42 U.S.C. § 2000e-2(k)(1)(A)(ii); *see also Albemarle*, 422 U.S. at 425. That necessarily involves a comparison of the current practice to the proposed alternative practice, both as to job relatedness and business necessity and as to adverse impact on minority applicants. Although the alternative practice need not measure the same characteristic as the challenged practice, it must be directed to the same business purpose, which need not encompass *all* aspects of effective

[22]Criterion-related validity is shown empirically by a correlation coefficient, expressed as a decimal value ranging from -1.0, a perfect inverse relationship, to 1.0, a perfect positive relationship.

job performance.[23] The alternatives must achieve the same business purpose, although they need not be the same type of tests as those challenged. Alternative employment practices, if the disparate impact analysis reaches the third step, are defined independently.[24]

B. Application

The United States challenges the use of the cognitive skills tests to eliminate from consideration for hire those applicants who score below certain cutoffs. Specifically, the United States identifies the following employment practices as unlawful: (1) initial screening using IPMA B-4, that is, using a cutoff score of 70%; (2) operational cut based on IPMA B-4 scores; (3) initial screening using ALERT, that is, using a cutoff score of 70%; (4) rank order hiring based on ALERT scores; (5) initial screening using ALARM, that is, using a cutoff score of 70%; (6) operational cut based on ALARM scores; and (7) rank order hiring based on ALARM scores.[25]

1. Initial Screening Using a Cutoff Score of 70%

---

[23]In *Dothard v. Rawlinson*, 433 U.S. 321 (1977), the Court rejected height and weight requirements for prison guards because the defendants failed to show any evidence correlating those requirements to the asserted business purpose – ensuring sufficient strength for adequate job performance. *Id.* at 331. The Court stated that measuring strength directly would be an acceptable alternative, *id.* at 332, although effective job performance obviously requires more than strength. There was no mention of replacing the height and weight requirements with tests that would address *other* requisites for effective performance. The Court infers that the proffered justification for the challenged practice, if demonstrated by a preponderance of the evidence, determines what must be shown as to any proposed alternative practices. That does not necessarily mean that an alternative practice cannot achieve *additional* business purposes. It does suggest, however, that any such additional benefits offered by the proposed alternative should not be counted for the "job relatedness/business necessity" component of the comparison of the current practice to the alternative, at least to the extent that such purposes were already achieved by the current process through components which the United States does not challenge.

[24]For example, even if the initial screening and operational cut were eventually found to comply with Title VII, the United States could propose a different method of selection for hiring from the eligibility list.

[25]The United States had also challenged rank order hiring based on IPMA B-4 scores. It, however, subsequently conceded that rank order hiring based on IPMA B-4 scores had no practical effect. The court, therefore, granted the City's Rule 52(c) motion with respect to this challenged practice. TR Vol. 2B at 55-56:17-19.

Three of the seven challenged practices deal with the initial screening of applicants using a cutoff score of 70% on the cognitive skills tests: practice number 1, initial screening using IPMA B-4, that is, using a cutoff score of 70%; practice number 3, initial screening using ALERT, that is, using a cutoff score of 70%; and practice number 5, initial screening using ALARM, that is, using a cutoff score of 70%.

a. IPMA B-4

i. Prima Facie Case

The court determines that the United States has established a prima facie case with respect to the use of a 70% score on the IPMA B-4 as an initial screen.[26] It is undisputed that the that the use of the IPMA B-4 with a 70% cutoff score resulted in statistically significant differences in passing rates among white, black and Hispanic test-takers. Specifically, the pass rates for minority test-takers on this test are statistically significantly lower than the pass rate for whites.

ii. Job Relatedness and Business Necessity

The burden then shifts to the City to establish that this practice is job-related and consistent with business necessity. *See Frazier*, 980 F.3d at 1525-26. To carry this burden, the City presented evidence regarding the development and validation of the IPMA B-4 and testimony from expert witnesses, Drs. Schemmer, Morris and Lundquist. Dr. Schemmer was one of the industrial organizational psychologists who developed the B-4 examination for IMPA. Dr. Schemmer testified that the developmental form of the B-4 examination contained 110 items and that the final version of the B-4 contained 85 items. He explained that the final version had fewer items than the

[26]In its proposed findings of fact and conclusions of law, the City admits that the United States has established its prima facie burden with respect to this challenged practice. Def. Proposed Findings of Fact and Conclusions of Law at 27.

developmental form because the number of items was reduced based on the data collected during the criterion-related validation phase.

Dr. Schemmer co-authored a report documenting the processes used for developing and validating the B-4 examination.[27] The development and validation of the B-4 examination included (1) a review of job analysis materials; (2) evaluation of the previous versions of the B-4 examination to identify items for replacement or revision; (3) development of a pool of items from which the replacements or revisions would be made; and (4) analysis of the examination for evidence of criterion-related validity. In short, Dr. Schemmer testified as to the validity of the B-4 examination and opined that the B-4 examination was "a valid and useful tool for the selection of entry-level firefighters." TR 3B 40-41:23-2.

Dr. Morris, an industrial organizational psychologist with Morris & McDaniel, testified that IPMAhired his firm in 1988 to revise some of its examinations, including the B-4 examination. Dr. Morris testified that based on the amount of money his firm was paid ($9,000) only minor revisions to the examination would have been made.[28] These revisions, Dr. Morris opined, would not have necessitated an additional validity study.

The United States does not directly dispute Dr. Schemmer's testimony regarding validity or job relatedness; instead, it contends that this testimony is not relevant because there is no evidence

---

[27]Dr. Schemmer's report is entitled, "Development and Validation of the B-3 and B-4 Entry-level Firefighter Examinations" ("Schemmer's Report"). At trial, Schemmer's Report (trial exhibit 692) was conditionally admitted into evidence. The court determines that sufficient evidence was subsequently presented to establish the relevance of the document, and therefore the exhibit is admitted into evidence.

[28]Dr. Schemmer testified that Morris & McDaniel would have had to substantially change at least half of the items on the examination *and* those items would have had to be dysfunctional in that they had no relationship to the job for the criterion-related validity evidence documented in his report to decrease to a questionable or unacceptable level. TR Vol. 3B at 44:19-24.

that the version of the B-4 examination that was administered by the City in 1992, namely, the 1989 version, was the same as the version that was the subject of Dr. Schemmer's validity study, namely, the 1986 version. Moreover, the United States contends that Dr. Morris's testimony that the changes his firm was hired to make to the 1986 version were not substantial is insufficient given that he has no recollection of the work his firm actually did.

The City, however, presented testimony by Neil Reichenberg, the Executive Director of the IPMA, disputing the existence of two versions (a 1986 version and a 1989 version) of the B-4 examination. Mr. Reichenberg has been employed with IPMA since 1980, first as its Director of Government Affairs and then as its Executive Director. Mr. Reichenberg testified that IPMA copyrights all of its examinations for security purposes and that the 1989 version of the B-4 examination is the only IPMA examination copyrighted with the "B-4" designation. Mr. Reichenberg's testimony supports the City's contention that there was only one version of the B-4 examination. Mr. Reichenberg further testified that IPMA re-designates examinations after substantial changes are made, which supports Dr. Morris's testimony that his firm did not substantially change the B-4 examination that had been the subject of Dr. Schemmer's study.

Finally, Dr. Lundquist testified that she performed two studies relating to the validity of the B-4 examination. In 1999, the first study focused primarily on content validity. To establish a link between the tasks required of an entry-level firefighter and the knowledge, skills, abilities and other (or personal) characteristics ("KSAOs")[29] required by the job, Dr. Lundquist relied in part on Dr.

---

[29]Dr. Landy explained that "knowledge" is the body of information relevant to the job which is acquired through training or study; "skill" is an act honed by practice; "ability" is an enduring characteristic, including mental abilities; and "other (or personal) characteristics" is personality or attitudinal variables such as dependability, initiative and honesty.

Wollack's job analysis.[30] She then established a link between the KSAOs and the test questions using a panel of industrial organizational psychologists. This link was further confirmed by the SMEs, as their answers generally supported the judgment of the panel. Dr. Lundquist also considered Dr. Schemmer's Report. Based on all of this information, Dr. Lundquist concluded that the B-4 examination was a valid selection device.[31]

The second study was conducted as a follow-up study to determine whether there were any significant changes to the firefighter job between 1986 (which was covered by Dr. Schemmer's Report) and 1992 and between 1992 and 1999 (which was covered by Dr. Wollack's job analysis). In this study, Dr. Lundquist used 21 SMEs, all of whom were currently working for the Fire Department and had been since 1992. The information gathered in the study revealed that the important tasks for a firefighter in the Fire Department had remained consistent over time.

Dr. Lundquist also analyzed the appropriateness of the 70% cutoff score using the Angoff Method.[32] The Angoff Method requires SMEs to estimate the probability that a minimally qualified individual would answer each individual test question correctly. The estimates are used to determine an overall passing score. The Angoff results revealed that if all estimates of the SMEs were used, the passing score was 68.84%. If all the estimates of the SMEs, excluding one outlier were used, the passing score was 71.5%. If the estimates of only the minority SMEs were used, the passing

---

[30]A job analysis is a method industrial organizational psychologists use to determine the tasks or duties of a job and the KSAOs related to successful performance of the job. Dr. Wollack's job analysis was verified by the information provided by approximately twenty incumbents from the Fire Department, who served as subject matter experts ("SMEs").

[31]Specifically, Dr. Lundquist testified that "it's my professional opinion that there was evidence of content, criterion, and construct related validity for the IPMAB-4." TR Vol. 4B at 12:5-9.

[32]Dr. Lundquist cited to professional literature that supports the use of the Angoff Method and followed the procedures for its implementation recommended in the literature.

score was 70.98%. Based on this data, Dr. Lundquist concluded that a passing score of 70% was reasonable.

The United States contends that the B-4 examination is not job-related and consistent with business necessity because the weights of the skill categories do not appropriately correspond with the importance of those skills for the performance in the job; it does not measure the attributes that are important for the performance of a firefighter; and it does not address any of the noncognitive aspects of job performance. There is no dispute that noncognitive skills are important and necessary for a firefighter and that the B-4 examination does not measure these skills. The evidence, however, establishes that the post-test components of the selection process *do* measure these skills. Thus, the court is not persuaded by this argument.

The United States further contends that Dr. Lundquist's use of the Angoff Method was flawed because of the disagreement and/or confusion among the SMEs about the link between the test items and the skills measured by those test items; and the use of two questions which were not appropriate for inclusion in an entry-level examination. In short, the United States contends that a 70% cutoff score was too high and thus results in a disparate impact. Dr. Lundquist, however, testified that the confusion among individuals lacking testing expertise is in inherent in this type of study, and therefore she also used a panel of industrial organizational psychologists. She further testified that even giving all test-takers credit for the two questions that should not have been included in the examination the cutoff score would be 68.85%.

Based on the testimony and evidence presented on this issue, the court determines that the City has carried it burden of establishing that the B-4 examination was job-related and consistent with business necessity. The United States must therefore establish that there are available

alternative practices that achieve the same business ends, with less racial impact. *See Frazier*, 980 F.3d at 1525-26.

iii. Alternatives

The only alternative practice identified by the United States is the ALARM examination. The City currently uses the ALARM examination. Indeed, Mr. Barnett, the City Civil Service Director, testified that the primary reason he replaced the B-4 examination with that of the ALARM examination was to reduce adverse impact. Even if the court agrees that the ALARM examination is a valid alternative to the B-4 examination, the United States also challenges the use of the ALARM examination in this litigation. The court finds it disingenuous to point to the ALARM examination as a valid alternative on one hand and portray it as unlawful on the other hand. Therefore, the court determines that the United States has failed to carry its burden of establishing the existence of a valid alternative.

b. ALERT and ALARM

i. Prima Facie Case

The court determines that the United States has established a prima facie case with respect to the use of a 70% score on the ALERT and ALARM examinations as an initial screen.[33] Specifically, the City concedes that the pass rates for black test-takers on these examinations are statistically significantly lower than the pass rate for whites. While the City contends that the United States's showing of a significant disparate impact against Hispanic test-takers was questionable, it concedes that ultimately it is of no consequence.

---

[33]Indeed, in its proposed findings of fact and conclusions of law, the City admits that the United States has established its prima facie burden with respect to this challenged practice. Def. Proposed Findings of Fact and Conclusions of Law at 36.

ii. Job Relatedness and Business Necessity

The burden then shifts to the City to establish that the practice of using these examinations is job-related and consistent with business necessity. *See Frazier*, 980 F.3d at 1525-26. To carry this burden, the City presented evidence from Dr. Wollack regarding the validation of the ALERT and ALARM examinations.[34] Dr. Wollack testified that he performed a validation study on each examination and concluded that both the ALERT and ALARM examinations measure abilities essential to the performance of the duties of police officers and firefighters, respectively; that the scores on the examinations are highly predictive of academy and state certification test performance; and that the 70% cutoff score on the examinations is reasonable. In short, Dr. Wollack testified that the ALERT and ALARM examinations are content valid.[35]

Specifically, Dr. Wollack testified that the SMEs for the ALERT examination identified reading, writing and interpersonal skills as essential skills for police officers, and the SMEs for the

---

[34]The ALERT examination was originally based on four multi-jurisdictional studies, including a study in Texas, Michigan, and California and an eight-state study. ALERT has also been the subject of 19 other validation studies involving 81 police departments, involving entry-level police officer positions. Similarly, since 1987, the ALARM examination has been the subject of 10 validation studies involving the entry-level firefighter position, including five content validity studies involving 67 fire departments.

[35]Dr. Wollack also presented evidence of criterion-related validity, which is not necessary to justify use of a content valid examination but rather serves as corroborative evidence of validity. Specifically, Dr. Wollack testified that his studies revealed a high correlation between scores on the ALERT and ALARM examinations and performance at the respective academies and state certification examinations: r = .77 (ALERT score and police academy performance); r = .71 (ALERT score and TCLEOSE examination); r = .76 (ALARM score and fire academy performance); r = .73 (ALARM score and TCFP examination). Indeed, Dr. Siskin, an expert witness presented by the United States, testified that the state certification examination scores were significantly correlated with the scores on the ALERT and ALARM examinations. The United States, however, contends that this correlation evidence is insufficient to support criterion-related validity because there is no evidence that a correlation exists between the ALERT and ALARM scores and any measure of on-the-job performance. The court notes that Dr. Wollack testified that the relationship between test scores and job performance is assumed to be linear. He further testified that he chose performance at the academies and on the state certification examinations because both represent an individual's ability to comprehend and apply the job knowledge of an entry-level police officer or firefighter.

ALARM examination identified reading skills as an essential skill for firefighters.[36] Dr. Wollack testified that these findings were consistent with the findings of other studies.[37] Additionally, Dr. Wollack's findings were corroborated by the testimony of the witnesses from the City's Police and Fire Departments.[38]

The court determines that the ALERT and ALARM examinations are job-related, even though they do not measure *all* of the cognitive skills or any of the noncognitive skills that are necessary for the performance of the duties of a police officer or firefighter, respectively. There is no dispute that police officers and firefighters must possess certain noncognitive skills to successfully perform their jobs. The evidence, however, establishes that the City's overall selection process measures the necessary cognitive and noncognitive skills of a police officer and firefighter, albeit not in the cognitive skills test portion of the selection process.

---

[36]Dr. Landy agreed that reading comprehension and writing ability are critical skills for a police officer, and comprehension, table interpretation and mechanical comprehension are necessary skills for a firefighter and thus measuring these cognitive skills during the selection process is appropriate.

[37]Dr. Jones, an expert witness called by the United States, agreed that better information regarding the job relatedness of an examination is yielded when data regarding the examination's effectiveness is gathered from various studies in various settings.

[38]The City presented four witnesses from the Police Department to testify regarding the cognitive demands of police officer training and the job itself: Officer Orman, who testified regarding the cognitive demands at the police academy; Officer Crockett, who testified regarding the cognitive demands of field training; Officer Hernandez, who testified regarding the cognitive demands of the daily performance of police officer duties; and Chief Wilson, who testified regarding the need for cognitive skills given the increased complexity of the duties of a police officer. Similarly, the City also presented four witnesses from the Fire Department to testify regarding the cognitive demands of firefighter training, including EMT and paramedic training, and the job itself: Battalion Chief Holcombe and Firefighter Balanciere, both of whom testified regarding the cognitive demands at the fire academy and during training; Firefighter Crews, who testified regarding the cognitive demands of paramedic training; and Chief Grammar, who testified regarding the cognitive demands that entry-level firefighters routinely perform in the exercise of their daily duties.

The United States challenges the City's practice of setting the cutoff score at 70%.[39] The City initially used a cutoff score of 70% on the ALERT and ALARM examinations because it believed it was required to do so by state law.[40] After this lawsuit was initiated, Dr. Wollack performed a two-step study to determine whether the 70% cutoff score on the ALERT and ALARM examinations was reasonable. First, incumbent supervisors in the Police Department estimated the percentage of entry-level police officers they had supervised whose reading and writing skills were sufficiently deficient as to impair their job performance.[41] Similarly, incumbent supervisors in the Fire Department estimated the percentage of entry-level firefighters they had supervised whose math, mechanical and table interpretation skills were sufficiently deficient as to impair their job performance.[42] Second, the ALERT and ALARM examinations were administered to 100 incumbent police officers and firefighters, respectively, to obtain normative score distributions. Dr. Wollack then combined this data with equivalent data previously gathered in larger jurisdictions where the same two-step process had been conducted. He explained that he combined the data because the sample size in Garland was too small to be reliable.

---

[39]Indeed, "the United States has not argued that 'the City's use of an initial cognitive screen is not a business necessity.' . . . To the contrary, the United States has repeatedly stated that use of a cognitive test as a screening device in the selection of police officers may be appropriate if the cutoff score used corresponds to the level of the measured skills needed for the job." Pl. Obj. at 22 (emphasis omitted).

[40]A state statute provides for an "entrance examination," which "must be based on the person's knowledge of and qualifications for fire fighting and work in the fire department or for police work and work in the police department and must inquire into the applicant's general education and mental ability." Tex. Local Gov't Code Ann. § 143.025(b). Applicants must score above 70% (as adjusted for veteran status) on the entrance examination to pass. *Id.* § 143.025(i).

[41]The estimate was 2%.

[42]The estimate was 4%.

The combined police supervisor sample yielded an 11% estimate of substandard performers which was applied to the larger normative score distribution, necessitating a cutoff score that would eliminate the bottom 11% of test-takers, namely, a cutoff score of 77%.[43] The same process was conducted with the fire data. The combined fire supervisor sample yielded a 10.5% estimate of substandard performers which was applied to the larger normative score distribution necessitating a cutoff score that would eliminate the bottom 10.5% of test-takers, namely, a cutoff score of 69%.[44] Based on his studies, Dr. Wollack concluded that a 70% cutoff score on both the ALERT and ALARM examinations is appropriate.

The United States contends that a cutoff score of 70% is too high, given that 27 police officer applicants and 16 firefighter applicants who scored below a 70% obtained employment elsewhere.[45] This alone, however, is simply insufficient to establish that 70% was an unreasonable cutoff score, as there could have been a number of reasons why these individuals were able to obtain employment elsewhere. By way of a nonexclusive list, these individuals might have been able to obtain employment elsewhere because of (1) the difficulty of the examination as compared to the City's; (2) a change in his or her study habits; (3) the testing environment on the day of the examination; and (4) other factors not related to the examination.

---

[43]Dr. Wollack testified that had he used the Garland police data only, it would have yielded a cutoff score of 70%.

[44]Dr. Wollack testified that his study used ALARM form 02, the most difficult of the alternate forms, and thus the recommended cutoff score would likely be slightly above 70%.

[45]The United States referred to these applicants as "false negatives." Of these false negatives, four of the police officer applicants testified at trial: Officer Hernandez, Officer Oliphant, Officer Gibson, and Officer Jemmott.

The United States also presented Dr. Landy who criticized Dr. Wollack for combining the Garland data with the data from the larger scale studies. The court notes, however, that had the data not been combined, the cutoff score would have been based on the scores of two police officers and four firefighters. Such a small sample could skew the results. Dr. Landy also criticized Dr. Wollack's methodology claiming that the question posed to the incumbents was susceptible to misinterpretation. There is no evidence, however, that any misinterpretation actually occurred. Dr. Landy acknowledged, however, that while he was in the public safety practice of his firm, the most common method of establishing a cutoff score was to administer the examination to a group of incumbents and then set the cutoff score at one standard deviation below the incumbent's mean score. Had this been done in this case, the cutoff score for ALERT would have been 76%, and the cutoff score for ALARM would have been 71%.

The United States also presented Dr. Siskin who concluded that, based on his study, ALERT scores have no value in predicting whether an applicant will successfully complete the police training academy, police field training or the first year on the job as a police officer.[46] Dr. Siskin also concluded, based on a regression analysis, that while there may be a correlation between ALERT scores and academy final grades, as Dr. Wollack testified, ALERT scores do not predict whether an applicant will pass or fail the police academy. Dr. Siskin testified that an applicant with an ALERT score of 60% is predicted to pass the academy, thus calling into question the necessity of a cutoff score of 70%. Dr. Siskin also analyzed the likelihood of passing the state certification examinations based on ALERT and ALARM scores. Dr. Siskin testified that his analysis revealed that applicants

---

[46]Dr. Siskin conceded, however, that his analysis confirmed Dr. Wollack's findings that ALERT and ALARM scores are significantly correlated with scores on the state certification examinations.

who scored at least 60% on the ALERT or ALARM examination and who later had the opportunity to take the state certification examination passed the state certification examination at a rate that was not significantly different than the pass rate of the applicants who were hired by the City.

In response, Dr. Wollack noted that Dr. Siskin's sample size did not include any person who either failed the ALERT examination or the training academy. Dr. Wollack testified that "if ALERT predicts a performance measure, and if the performance is job related, then ALERT predicts job performance." TR Vol. 5A at 60:13-15. The same would hold true for ALARM. Dr. Wollack testified that he performed an analysis similar to that conducted by Dr. Siskin but used more complete data gathered from another jurisdiction and concluded that there is "a very high statistical correlation between ALERT scores and passing the training academy." TR Vol. 5A at 61:7-8.

Moreover, two expert witnesses called by the United States, Drs. Landy and Jones, acknowledged that there is *no* single acceptable method for establishing a cutoff score on an employment examination. Dr. Jones testified that all methods of establishing a cutoff score are based on arbitrary decisions and therefore the cutoff score need only be practical and sensible and based on good business judgment given the available data.

Based on the testimony and evidence presented on this issue, the court determines that the City has carried it burden of establishing that the ALERT and ALARM examinations are job-related and consistent with business necessity. The United States must therefore establish that there are available alternative practices that achieve the same business ends, with less racial impact. *See Frazier,* 980 F.2d at 1525-26.

iii. Alternatives

As an alternative to the ALERT examination, the United States proposes the Protective Services Profile ("PSP"), marketed by Aon Corporation. As an alternative to the ALARM examination, the United States proposes the Selection Solutions Firefighter Examination, published by CWH Management, Inc. ("CWH examination").

(1) PSP

The PSP is a written, multiple choice examination consisting of four components: a cognitive component called the "Following Policies and Procedures" ("FPP");[47] a noncognitive biodata component called the "Work and School Experience Form" ("WSE");[48] and two personality tests called the "Survey of Life Experiences-Adjustment" ("SOLE-Adjustment")[49] and "Survey of Life Experiences-Dependability" ("SOLE-Dependability").[50] The PSP is broader in content coverage than ALERT, as it measures both cognitive and noncognitive skills.[51] The total score of the PSP is calculated in a compensatory manner, by equally weighing the four components of the examination

---

[47]The FPP is a multiple-choice cognitive ability test that measures reading comprehension, reasoning, and the ability to make decisions based upon information provided in reading passages.

[48]The WSE is intended to measure on-the-job performance by scoring the applicant's ability to differentiate between the top performing and bottom performing police officers.

[49]The SOLE-Adjustment is a 64-item personality scale focused on identifying individuals who are even-tempered, think clearly, maintain their composure under stress and are emotionally stable.

[50]The SOLE-Dependability is an 82-item personality scale focused on identifying individuals who are reliable, well-organized, honest and trustworthy.

[51]The Aon PSP was developed based on the findings gathered in several studies, including studies in Nassau County, New York; Louisiana; Parma, Ohio; and Arkansas. The PSP used in Louisiana and Parma was not the same as the PSP used in Arkansas, and the PSP detailed in Dr. Jones's expert report is not the same as the PSP used in Louisiana, Parma or Arkansas.

such that a higher score on one component can compensate for a lower score on another component.[52]

Dr. Jones, who was the project director for the development and validation of the PSP as an entry-level police officer selection device, testified that the PSP would be at least equally valid as the ALERT examination; however, he had not seen the ALERT examination at the time he made this conclusion. Further, although critical of the use of ALERT as an initial screen, Dr. Jones recommends that the cognitive component of the PSP be used as a screening device with a separate cutoff score, thus precluding the applicants who fail the initial cognitive screen from providing information regarding their noncognitive attributes.[53] Dr. Jones acknowledged, however, that in at least one of his studies, the use of the cognitive portion of the PSP as an initial screen resulted in an adverse impact against black test-takers.

Dr. Jones also testified that the correlation between the cognitive component of the PSP and overall job performance is statistically significant, but that combining noncognitive components with the cognitive component increases the validity of the PSP.[54] He further stated that the correlation between all the components of the PSP to job performance cannot be directly compared to the correlation between ALERT and job performance because there is no study comparing the criterion-related validity of ALERT to job performance. Dr. Jones further testified that, with respect to

---

[52]The test is compensatory only to the extent that the applicant passes the cognitive skills component. In other words, applicants who pass the cognitive skills component are ranked based on the total score of all the components of the PSP.

[53]Aon acknowledged that "[a] strong 'business necessity' argument can be made that the cognitive test should be used to screen out the small number of applicants at very low cognitive ability levels who have little realistic prospect of either completing training, or emerging from training as a well qualified officer." Pl. Ex. 355 at JEF 448.

[54]The correlation between all the components of the PSP and job performance is .62.

criterion-related validity, the PSP's cognitive component "produced validity coefficients not as great as but close to the validity coefficients produced for the ALERT in predicting academy performance." TR Vol. 7B at 25:7-10.

Dr. Jones compared the validity of the PSP used in Louisiana to the validity of a cognitive-only examination. Based on these findings, he concluded that the PSP demonstrated a substantially higher level of validity to job performance than a cognitive-only examination. Although the PSP used in this study is different than the PSP which the United States proposes to use in Garland, Dr. Jones testified that he would expect only a modest change to the validity of the PSP. Dr. Jones also examined the adverse impact of the PSP as compared to ALERT. Dr. Jones opined that even with a cutoff score on the cognitive component, the PSP would allow more blacks and Hispanics to continue through the selection process. Dr. Jones's comparison, however, was based on Dr. Siskin's "effective cutoff score" analysis of ALERT, *see infra.*, not upon the 70% cutoff score used by the City as an initial screen. The court is thus not persuaded by Dr. Jones's testimony.

Further, the evidence establishes that there is some overlap between the noncognitive skills evaluated in the PSP and the post-test components of the City's selection process. Therefore, comparing the PSP, a "comprehensive/compensatory" examination, to the initial screening would be at least partially redundant with other components of the current overall selection process. For all the reasons herein, the court determines that the United States failed to demonstrate that the PSP is a legally appropriate alternative, that is, that it would equally serve the City's legitimate business needs but with less adverse impact resulting from its use of ALARM.

(2) CWH Examination

Next, the United States proposes the CWH examination as an alternative to ALARM. The United States presented Dr. Hornick. Dr. Hornick testified regarding the development and validation of the CWH examination. Specifically, Dr. Hornick performed a criterion-related validation study; he did not perform a content or construct validation study.[55]

The CWH examination is a multiple-choice examination, which includes 156 questions of which 136 questions are scored. The 136 questions are distributed among four components: cognitive skills, interpersonal skills, self-awareness skills, and practical skills.[56] The relative influence of each of the four components on the overall composite score is affected by the amount of unique variance contributed by each component, as well as the common variance shared by the various components ("rational weighting approach"). The CWH examination is typically used as the first step in the selection process; it does not replace the need for a physical agility test, oral interview, or background investigation.

The criterion-related validation study of the CWH examination involved 17 fire departments and 769 incumbent firefighters. Garland was not one of the fire departments; in fact, none of the fire departments studied was located in Texas. The criterion used was job performance, as evaluated by each incumbent firefighter's supervisors, peers and self.[57] In short, Dr. Hornick testified that the criterion-related validity of the CWH examination is .27 (uncorrected).

---

[55]Dr. Barrett opined that the CWH examination was not content or construct valid.

[56]Noncognitive items are used to examine an applicant's self-awareness and interpersonal skills; Dr. Hornick conceded that some of the noncognitive items can be faked. Dr. Barrett opined that some of the noncognitive items do not relate to performance of the entry-level firefighter job.

[57]Dr. Barrett testified that self-ratings and peer ratings are unreliable measures of job performance.

The court, however, finds Dr. Hornick's criterion-related validation study suspect because he eliminated data from nearly 40% of the study participants,[58] and excluded other individuals as "statistical outliers." Further, the evidence established that CWH's financial condition affected the quality of the validation study. For example, Dr. Hornick admitted that his job analysis suffered due to the financial condition of CWH in that he was unable to "devote face-to-face time meeting with the raters and reviews to have them go through the process." TR Vol. 9A at 68:4-23.

Dr. Hornick also testified regarding the adverse impact of the CWH examination. He found that the standardized mean score difference for black test-takers as compared to white test-takers on the CWH examination (.43 standard deviations) is below the black/white standardized mean score difference on the ALARM (1.30 standard deviations).[59] He testified that the same is true for Hispanic and white test-takers on the CWH examination (.14 standard deviations) as compared to the ALARM (.73 standard deviations). Moreover, Dr. Hornick testified that as measured by the 80% rule, the CWH examination has no adverse impact against Hispanics at any cutoff score and less adverse impact against blacks than the use of the ALARM. Dr. Hornick conceded, however, that at the time of his expert report, he had never seen the ALARM examination.

Dr. Hornick did not analyze the pass rates for whites and minorities on the cognitive component of the CWH because he does not recommend that a cutoff score be used on the cognitive component. Dr. Hornick does not recommend that the City use only the cognitive component of the

---

[58]Dr. Hornick explained that he did not have complete data from all three rating sources (supervisor, peer and self) for approximately 300 incumbents, and therefore he did not use these incumbents in the validity study. Dr. Hornick maintained that the removal of incumbents with incomplete data did not affect the validity calculation.

[59]Dr. Siskin testified that if a test has a lower mean score difference, it has less adverse impact.

CWH examination, but rather that the City adopt the entire examination. The court notes, however, that ALARM is a purely cognitive examination. Moreover, Dr. Hornick strongly recommends that a validation study be conducted if Garland were to use the CWH examination, TR Vol. 9A at 34:6-8, a test he developed and for which he stands to receive substantial income if Garland were to use the CWH examination and contract with CWH to conduct the necessary validation study, TR Vol. 9A at 34:6-8.

The court further determines that Dr. Hornick's conclusion "there are *no* other selection devices for the job of firefighter that [he thinks] are good or would recommend" is disingenuous, self-serving, and exaggerated. TR Vol. 9A at 72-73:24-1 (emphasis added). The court is not persuaded by Dr. Hornick's testimony.

Finally, the evidence establishes that there is some overlap between the noncognitive skills evaluated in the CWH examination and the post-test components of the City's selection process. Therefore, comparing the CWH examination, a "comprehensive/compensatory" examination, to the initial screening use of ALARM would be at least partially redundant with other components of the current overall selection process. For all the reasons stated herein, the court, therefore, concludes that the United States failed to demonstrate that the CWH examination would equally serve the City's legitimate business needs with less adverse impact resulting from its use of ALARM.

2. Operational Cut Based on Cognitive Skills Tests Scores[60]

Two of the seven challenged practices deal with the rank order processing of applicants: practice number 2, operational cut based on IPMA B-4 scores; and practice number 6, operational cut based on ALARM scores.

a. IPMA B-4

The United States contends that the City's use of an "operational cutoff score" on the IPMA B-4 examination adversely impacted black and Hispanic test-takers. As support, the United States presented the testimony of Dr. Siskin. Dr. Siskin testified that he performed a mean difference analyses, that is, he analyzed the differences in mean B-4 examination scores obtained by blacks and Hispanics as compared to whites. Dr. Siskin explained that under the common assumption that scores are normally distributed, the score distribution curve for minority applicants is "shifted to the left" as compared to the score distribution for white applicants when the mean scores are significantly lower for minority applicants than for white applicants. He testified that when the score distribution of minority applicants is shifted to the left, there will be a disparate impact against minority applicants at any cutoff score, and that the farther the curve is shifted to the left the greater the impact.

Dr. Siskin tested whether the minority score distribution curves for the B-4 examination were shifted to the left to a statistically significant degree. He concluded that the differences were statistically significant and were not attributable to chance. Specifically, the difference in mean

---

[60]The United States refers to the lowest score of an applicant who was invited to proceed to the physical agility test as an "operational cutoff score." The City contends that it does not have "operational cutoff scores," and refers to the employment practice at issue as "rank order processing" based on the cognitive skills examination scores. The court determines that the label used – "operational cutoff score" or "rank order processing" – is a difference with no distinction and will use the labels interchangeably.

scores between black applicants and white applicants was equivalent to 11.74 units of standard deviation, and the difference in mean scores between Hispanic applicants and white applicants was equivalent to 5.25 units of standard deviation.

The court, however, finds Dr. Siskin's analysis suspect for two reasons. First, data regarding the number of test-takers who advanced to the physical agility test was not addressed in the parties' stipulated database. Dr. Siskin instead relied on City documents that appeared to him to be partial eligibility lists of test-passers. These lists were marked with handwritten notes and shorthand references from which he concluded what occurred to the applicants during the testing process. Dr. Siskin could not identify the author of the documents or whether they were compiled contemporaneously with the administration of the selection process by a person with knowledge of those processes, and no witness or other evidence verified the accuracy of the documents. The City, however, presented evidence that these documents contained material inaccuracies.

Second, Dr. Siskin did not use the population to whom the "operational cutoff score" would have applied. He instead based his mean score analysis on all test-takers, whether they passed or failed the B-4 examination.[61] Dr. Siskin explained his use of the test-taker population by stating that given the "shift to the left," disparate impact will occur at any given cutoff score. Dr. Stoikov, however, compared the number of minority and white test-passers with the number of minority and white test-passers who passed the "operational cutoff score," that is, those who were invited for further processing. She found a shortfall of .1 blacks, which she concluded was not significant, and no significant impact against Hispanics.

---

[61] Dr. Stoikov described Dr. Siskin's study as a "proxy study" because it was not applied to the exact population affected by the challenged practice.

Dr. Siskin also performed an "effective cutoff score"[62] analysis. Dr. Siskin's "effective cutoff score" analysis purports to calculate the shortfalls[63] in minorities who scored at or above the "effective cutoff score" from among all test-passers to measure the magnitude of the disparity caused by the rank-order processing. The United States contends that because the "effective cutoff score" equaled the "operational cutoff score"on the B-4 examination,[64] Dr. Siskin's "effective cutoff score" analysis is useful in analyzing the "operational cutoff score." In short, Dr. Siskin concluded that the "operational cutoff score" had a disparate impact on minorities. Specifically, he testified that of the 41 blacks, 24 Hispanics, and 346 whites who took the B-4 examination in 1992, no minorities and 27 whites scored at or above the "operational cutoff score." Thus, the operational pass ratio for minorities was 0% and for whites was 7.8%. As with the mean difference analysis, however, Dr. Siskin did not use the population to whom the "operational cutoff score" would have applied, that is, applicants who passed the cognitive skills test.

The court finds Dr. Siskin's testimony unpersuasive. Accordingly, the court determines that the United States has failed to establish a prima facie showing the City's practice of "rank order

---

[62]The United States refers to the total score (including veteran preference points, if any) of the lowest-scoring applicant hired from the eligibility list as the "effective cutoff score." The City contends that it never had a practice of using an "effective cutoff score" and that the term was created by Dr. Siskin. The City contends that the score that the United States refers to as the "effective cutoff score" merely evidences that there were no eligible applicants (because they failed post-test components of the selection process) with scores below the score of the lowest-scoring applicant that was hired.

[63]Dr. Siskin described a shortfall as the additional number of minority applicants who would have scored at or above the "effective cutoff score" if minority and white applicants were equally distributed among the ranks of test-passers.

[64]The United States contends that in 1992, the effective and operational cutoff scores on the B-4 examination were the same, namely, 89.41, because the lowest-scoring applicant who was allowed to take the physical agility test was also the lowest-scoring applicant hired from the eligibility list.

processing" based on the applicant's score on the B-4 examination resulted in an adverse impact against any group.[65]

b. ALARM

The United States contends that the City's use of an "operational cutoff score" on the ALARM examination adversely impacted black and Hispanic test-takers in the years 1994 through 1998.[66] As support, the United States presented the testimony of Dr. Siskin.

As with his analysis of the "operational cutoff score" on the B-4 examination, Dr. Siskin compared the percentage of minority test-takers who scored at or above the "operational cutoff score" on the ALARM to the percentage of similarly situated white test-takers. Also as with his

---

[65]Even if the United States has established a prima facie showing, which it has not, the City presented sufficient evidence to establish a business necessity and job relatedness for its practice of "rank order processing." Specifically, the City presented Chief Grammar and Captain Tokar who testified that it is impractical to send every applicant who passes the cognitive skills examination to the physical agility test, due to economic, personnel and time constraints. Captain Tokar testified that the Fire Department spends approximately $17,000 to process a group of 30 to 50 applicants through the post-test selection process. The United States presented no evidence to dispute the necessity of "rank order processing" as a cost-savings measure. The United States contends that the City presented no evidence that the "operational cutoff score" of 89.41% is job-related and consistent with business necessity. The City responds, however, that it does not have a practice of using "operational cutoff scores." The evidence established that applicants who pass the cognitive skills test but rank below the top 30 to 50 scorers are not eliminated from the hiring process; instead, they remain eligible for hire until the eligibility list expires. The City also points to Dr. Schemmer's testimony that the B-4 examination is a typical cognitive skills or abilities test which "provide[s] amongst the most sufficient and effective mechanisms for the rapid identification and [at an] early point in the selection process of those individuals relatively more likely to succeed in training and, in turn, exhibit successful job performance" (TR Vol. 3A at 187:20-25) as support for the job relatedness of "rank order processing." The court notes that the United States did not identify or present any evidence of an alternative to the City's use of "rank order processing," that is, another method of selecting from among the applicants who passed the B-4 examination. To the extent that the United States contends that the CWH examination is a lawful alternative to the B-4 examination, the court disagrees. The court incorporates by reference as if fully stated herein its findings and conclusions regarding the CWH examination. *See supra.*

[66]The United States contends that the "effective cutoff scores" and "operational cutoff scores" on the ALARM examinations are as follows: in 1994, both scores were the same at 90%; in 1995, an effective cutoff score of 95% and an operational cutoff score of 91%; in 1996, an effective cutoff score of 94% and an operational cutoff score of 91%; in 1997, an effective cutoff score of 92% and an operational cutoff score of 89%; and in 1998, both scores were the same at 91%.

analysis of the "operational cutoff score" on the B-4 examination, Dr. Siskin did not test the statistical significance of the disparity in minority and white pass rates and instead calculated the 80% rule ratio of black/white and Hispanic/white pass rates at the "operational cutoff score." Dr. Siskin calculated that over the entire period of 1994 through 1998, the black/white operational pass ratio was 13.9% and that the Hispanic/white ratio as 44.4%. Dr. Siskin compared these ratios to the black/white and Hispanic/white pass ratios at the 70% cutoff score, 50.1% and 76.2%, respectively, and concluded that use of the "operational cutoff score" had a disparate impact against minority applicants over and above the disparate impact resulting from the 70% cutoff score. The court, however, finds Dr. Siskin's testimony unpersuasive for the same reasons discussed with regard to the "operational cutoff score" analysis of the B-4 examination, namely, Dr. Siskin's use of unauthenticated eligibility lists[67] and use of the wrong population, that is, all test-takers.

Dr. Siskin also performed an analysis of the impact of the "operational cutoff scores" on test-*passers*. Dr. Siskin testified that for the period from 1994 through 1998, there is a statistically significant disparity (equivalent to 4.30 units of standard deviation) among test-passers between black/white operational pass rates. Similarly, Dr. Siskin testified that for the period from 1994 through 1998, there is a statistically significant disparity (equivalent to 2.37 units of standard deviation) among test-passers between Hispanic/white operational pass rates. Additionally, Dr. Siskin testified that during the same time period, the 80% rule was not met for minority test-passers (23.93% for blacks and 55.72% for Hispanics).

Dr. Stoikov pointed out that Dr. Siskin's analysis is not based on the data from the stipulated database. Dr. Stoikov reran Dr. Siskin's analysis using data from the stipulated database. She found

---

[67]Dr. Siskin also relied on Requests for Admissions that were later withdrawn by stipulation.

that a disparity equivalent to 3.42 units of standard deviation existed between black/white operational pass rates, and a disparity equivalent to 1.97 standard deviations existed between Hispanic/white operational pass rates.[68] With respect to the disparity between Hispanics and whites, she testified that the difference is on the borderline of statistical significance and much lower than that reported by Dr. Siskin. Moreover, Dr. Stoikov testified that she too was not using the correct population, that is, the operational passers who were *invited* to take the physical agility test, not just those who actually took it, because the data documenting the correct population is unavailable.

Dr. Stoikov further testified that, given that rank-order processing occurred on a yearly basis, a year-by-year analysis will have more "relation to reality" than an aggregation. Dr. Stoikov reran the operational pass analysis on a yearly basis. She found a shortfall of black operational passers that fall into the intermediate or questionable are of statistical significance in 1997 and 1998, and no significant shortfall in black operational passers in any other year. She further found no year in which there was any significant shortfall of Hispanic operational passers, and in fact, she found that in 1997, Hispanics were overrepresented among operational passers.

The court finds Dr. Stoikov's testimony more persuasive than that of Dr. Siskin. Accordingly, the court determines that the United States has failed to establish a prima facie showing the City's practice of "rank order processing" based on the applicant's score on the ALARM examination resulted in an adverse impact against any group.[69]

---

[68]In terms of number of people, the disparity is equivalent to less than one-tenth of one Hispanic operational passer.

[69]Even if the United States has established a prima facie showing, which it has not, the City presented sufficient evidence to establish a business necessity and job relatedness for its practice of "rank order processing." Specifically, the City's evidence regarding the cost-savings associated with "rank order processing" and its contention that it does not have a practice of using "operational cutoff scores"as detailed

3. Rank Order Hiring Based on Cognitive Skills Tests Scores

The United States contends that the City's practice of hiring individuals from the final eligibility list in descending rank-order based on their ALERT or ALARM scores results in a disparate impact against blacks and Hispanics.[70] In support, the United States presented Dr. Siskin. Dr. Siskin testified that he performed several analyses and concluded that there is a statistically significant relationship between the applicant's ranks on those examinations and their race or national origin. In other words, when applicants are ranked based on their scores on the cognitive tests, there is a difference between the ranks of whites and minorities that is real and not attributable to chance.

In his analyses, Dr. Siskin used the "effective cutoff scores" of the ALERT and ALARM examinations. He acknowledged that there could be no adverse impact from rank-order hiring in the years that the City exhausted the eligibility lists. The parties stipulated that there is no year during the relevant time period that the Fire Department exhausted its eligibility lists. The evidence, however, established that the Police Department exhausted its eligibility list in 1992, 1995, and 1997. Therefore, the court determines that there can be no disparate impact from rank-order hiring based on ALERT scores in 1992, 1995, and 1997.

---

in the section addressing "rank order processing" and the B-4 examination, *see supra.*, are also applicable here. The City also relied on the testimony of Dr. Wollack regarding the correlation between the ALARM scores and academy performance and scores on the state certification examination as support for the proposition that "rank order processing" helps avoid the costs of processing and training replacement recruits. The court notes that the United States did not identify or present any evidence of an alternative to the City's use of "rank order processing," that is, another method of selecting from among the applicants who passed the ALARM examination. To the extent that the United States contends that the CWH examination is a lawful alternative to the ALARM examination, the court disagrees. The court incorporates by reference as if fully stated herein its findings and conclusions regarding the CWH examination. *See supra.*

[70]Two of the seven challenged practices deal with rank order hiring: practice number 4, rank order hiring based on ALERT scores; and practice number 7, rank order hiring based on ALARM scores.

With respect to the Police Department eligibility lists in 1996 and 1998, Dr. Siskin opined that the lists were not exhausted, even though there were no candidates who scored below the "effective cutoff score" left on the list, that is, there were no eligible applicants who were not hired. Dr. Siskin explained that based on the deposition testimony of Chief Wilson, he concluded that the City would not have hired additional applicants had there been additional applicants on the eligibility list. At trial, however, Chief Wilson testified that the Police Department would have hired additional applicants in 1996 and 1998 had there been any on the eligibility list. The court therefore finds that Dr. Siskin's conclusion that the City would not have hired additional applicants had there been any was incorrect. The court further finds that the Police Department eligibility lists for 1996 and 1998 were exhausted. Accordingly, the court determines that there can be no disparate impact from rank-order hiring based on ALERT scores in 1996 and 1998.

The court next considers whether rank-order hiring resulted in a disparate impact in the Police Department in 1993 and 1994 and in the Fire Department from 1993 through 1998. Dr. Siskin concluded, using the rank sum test,[71] that there is evidence that minorities who pass the cognitive examination are "more likely to be at [a] lower rank proportionately than whites" on the eligibility list. TR Vol. 2B at 17:6-16. Dr. Siskin performed the rank sum test on the eligibility list of all test-passers. In other words, he used the eligibility lists created immediately following the administration of the cognitive skills test, as opposed to the eligibility lists created after the successful completion of the selection process ("final eligibility list").[72]

---

[71]The rank sum test is used to analyze whether there is a statistically significant relationship between race or national origin and rank-order.

[72]The City refers to the list from which the applicants were ultimately hired as the "final eligibility list."

The City contends that Dr. Siskin used the wrong population because rank-order hiring is only applied to applicants who not only pass the cognitive skills test but also the post-test components. In other words, the City contends that the correct population was the applicants who were on the "final eligibility list." Dr. Siskin explained that the eligibility lists he used are appropriate because (1) the "final eligibility list" is a much smaller sample than the eligibility list he used which could result in an erroneous conclusion; and (2) the relative ranks of the applicants do not change after they pass the cognitive skills examination, as the remaining components of the selection process are pass-fail. Dr. Stoikov conceded that an analysis of the test-passers is not "far off" from that of those applicants who were ultimately hired, assuming the post-test components of the selection process do not advantage or disadvantage minorities. She further explained, and Dr. Siskin conceded, that the actual rank of the applicants can change, as the lists narrows and the spread lessens. Dr. Stoikov also reiterated her criticisms of Dr. Siskin's analysis, namely, that it was based on data not included in the stipulated database, and it improperly aggregated ranks across years.

Dr. Siskin also performed a mean difference analysis and concluded that the minority score distribution curves were shifted to the left to a statistically significant degree and the differences in scores were not attributable to chance. In other words, differences in mean scores for all test-takers indicate that minorities will be clustered at the bottom, or ranked proportionately lower than whites, among the test-passers.

The City again argued, and Dr. Stoikov agreed, that Dr. Siskin used the wrong population. Dr. Stoikov testified that Dr. Siskin's analysis was improper because he considered all test-takers. Dr. Siskin, however, explained that given the "shift to the left" of the score distributions for minority applicants, disparate impact will occur at any cutoff score, including the "effective cutoff score."

Dr. Siskin further testified that it is important to analyze the entire sample of test-takers because analyzing only the test-passers can result in erroneous conclusions and would not isolate the impact caused by the challenged practices from the unchallenged practices. Dr. Siskin conceded, however, that he never attempted to separate his analysis of rank-order hiring from his analysis of all test-takers. Since he failed to even attempt to do so, his opinion is of minimal value.

Lastly, Dr. Siskin calculated shortfalls in minorities who scored at or above the "effective cutoff score" from among all test-passers to measure the magnitude of the disparity caused by the rank-order hiring. Dr. Siskin explained that in order for there to be a shortfall of minority effective-passers, they would be clustered in the lower ranks of the test-passers; thus, he measured the effect of minorities being clustered at the bottom of the group of test-passers.

Dr. Siskin concluded that there is an actual disparity caused by the City's rank-order hiring based on the ALERT and ALARM scores. Specifically, he concluded that in the Police Department, for the aggregated period from 1992 through 1995, there was a shortfall of two black and four Hispanic expected effective passers among the test-passers.[73] Similarly, Dr. Siskin concluded that for the aggregated period from 1994 through 1998, there were shortfalls of 11 black and 8 Hispanic expected effective passers among the test-passers.

The court finds, however, that Dr. Siskin's "effective cutoff score" analysis conflates rank-order hiring with the disqualifying effects of the post-test components. Specifically, Dr. Siskin testified that he considered an applicant who successfully completed the hiring process, including

---

[73]Dr. Siskin also found that from 1996 through 1998, there was a shortfall of 1 black (in 1996), but he would not have been placed on the final eligibility list because he failed some of the post-test components, and no shortfall of Hispanics. Dr. Siskin's findings further support the court's determination that the Police Department's rank-order hiring did not result in a disparate impact from 1996 through 1998. *See supra.*

passing the cognitive skills examination, but was not ultimately hired as effectively failing the cognitive skills examination. Similarly, he explained that he considered an applicant who passed the cognitive skills examination but failed a post-test component of the hiring process as effectively failing the cognitive skills examination. Dr. Siskin admitted that for purposes of this analysis the actual reason the applicant was not hired is of no moment.

Dr. Stoikov analyzed the practical significance or the magnitude of the effect of rank-order hiring. She found that the percentage of minority and white test-passers who were hired are not significantly different, and thus concluded that ranking could not have adversely affected the chances of being hired. Dr. Stoikov also analyzed the hiring consequences of the City's use of ALERT and ALARM. She concluded that in the Police Department, there was no significant shortfall in expected Hispanic hires, and one additional black hire would have eliminated the significant shortfall in expected black hires. She further concluded that in the Fire Department from 1992 through 1998, there was no significant shortfall in expected Hispanic hires, and a significant shortfall of 1.2 black hires.

Finally, Dr. Stoikov compared test-passers to effective passers in the Fire Department using Dr. Siskin's data and effective cutoff scores. Dr. Stoikov testified that while minority test-passers were hired at a lower rate than white test-passers, the difference is not statistically significant. She found that there was no shortfall in black effective passers large enough to reach the gray area of statistical significance, except in 1998, and that there was no shortfall in Hispanic effective passers in either a yearly analysis or in the aggregate. The court finds Dr. Stoikov's testimony persuasive.

For the reasons herein stated, the court finds Dr. Siskin's analyses is less persuasive than those of Dr. Stoikov. Accordingly, the court determines that the United States has failed to establish

a prima facie showing the City's practice of "rank order hiring" based on the applicant's score on the ALERT or ALARM examination resulted in an adverse impact against any group.[74]

III. Conclusion

For the reasons stated herein, the court concludes that the United States failed to prove by a preponderance of the evidence that any of the following employment practices used by the City violated Title VII: (1) initial screening using IPMA B-4, that is, using cutoff score of 70%; (2) operational cut based on IPMA B-4 scores; (3) initial screening using ALERT, that is, using a cutoff score of 70%; (4) rank order hiring based on ALERT scores; (5) initial screening using ALARM, that is, using a cutoff score of 70%; (6) operational cut based on ALARM scores; and (7) rank order hiring based on ALARM scores. Accordingly, the City is entitled to judgment in its favor, and all claims brought by the United States are dismissed with prejudice. The court will render judgment in favor of Defendant City of Garland by separate document as required by Fed. R. Civ. P. 58.

---

[74]Even if the United States has established a prima facie showing, which it has not, the City presented sufficient evidence to establish a business necessity and job relatedness for its practice of "rank order hiring." Specifically, Dr. Wollack testified that rank-order hiring based on ALERT and ALARM scores is psychometrically appropriate because there was an adequate job analysis, the examinations were reliable, and there was a useful spread of scores. Dr. Wollack further testified that rank-order hiring is appropriate because the relationship between test scores and job performance is assumed to be linear. The court incorporates by reference its discussion of Dr. Wollack's testimony regarding the validity of the ALERT and ALARM examinations. *See supra*. As an alternative for rank-order hiring based on the ALERT and ALARM scores, the United States proposes rank-order hiring based on the scores of the PSP (for police) and CWH examination (for fire). In other words, the United States proposes that the City use the PSP and CWH examinations as initial screens and after successful completion of the unchallenged post-test components of the hiring process, the City could rank-order hire based on the scores of these examinations. The court, however, has already found that the use of ALERT and ALARM with a cutoff score of 70% is lawful. Thus, to be an alternative to the City's current practice of rank-order hiring, the City would have to hire applicants in rank-order based on their scores on PSP or CWH examinations which was administered after these applicants successfully passed the initial screen, namely, the ALERT or ALARM, and the post-test components. The United States does not identify this as an alternative. To the extent that the United States contends that the PSP or CWH examination are lawful alternatives, the court disagrees. The court incorporates by reference as if fully stated herein its findings and conclusions regarding the PSP and CWH examinations. *See supra.*

It is so ordered this 31st day of March, 2004.

Sam A. Lindsay
United States District Judge