

UNITED STATES DISTRICT COURT  
EASTERN DISTRICT OF NEW YORK

-----X  
UNITED STATES OF AMERICA,

Plaintiff,

-and-

THE VULCAN SOCIETY, INC., MARCUS  
HAYWOOD, CANDIDO NUÑEZ,  
ROGER GREGG,

Plaintiff-Intervenors,

-against-

THE CITY OF NEW YORK, FIRE  
DEPARTMENT OF THE CITY OF NEW  
YORK, NEW YORK CITY DEPARTMENT  
OF CITYWIDE ADMINISTRATIVE  
SERVICES, and MAYOR MICHAEL  
BLOOMBERG and NEW YORK CITY FIRE  
COMMISSIONER NICHOLAS SCOPPETTA,  
in their individual and official capacities,

Defendants.

-----X  
NICHOLAS G. GARAUFIS, United States District Judge.

From 1999 to 2007, the New York City Fire Department used written examinations with discriminatory effects and little relationship to the job of a firefighter to select more than 5,300 candidates for admission to the New York City Fire Academy. These examinations unfairly excluded hundreds of qualified people of color from the opportunity to serve as New York City firefighters. Today, the court holds that New York City's reliance on these examinations constitutes employment discrimination in violation of Title VII of the Civil Rights Act of 1964.

## I. INTRODUCTION

In recent years, black and Hispanic residents of New York City (the “City”) have come to comprise a substantial portion of the City’s population, but their representation in the New York City Fire Department (“FDNY”) has remained extraordinarily low.<sup>1</sup> In 2002, the New York City Department of City Planning identified 25% of the City’s residents as black and 27% of its residents as Hispanic.<sup>2</sup> At the same time, however, only 2.6% of its firefighters were black and 3.7% of its firefighters were Hispanic.<sup>3</sup> When this litigation commenced in 2007, the percentages of black and Hispanic firefighters had increased to just 3.4% and 6.7%, respectively.<sup>4</sup> In other words, on a force of 8,998 firefighters, there were just 303 black firefighters and 605 Hispanic firefighters. These numbers stand in stark contrast to some of the nation’s other large cities, such as Los Angeles, Chicago, Philadelphia, and Houston, where minority firefighters have been represented in significantly higher percentages.<sup>5</sup>

In this case, Plaintiff United States of America (the “Federal Government”) as well as the Vulcan Society, Inc., Marcus Haywood, Candido Nuñez and Roger Gregg (the “Intervenors”), have sued to enforce the right of black and Hispanic candidates to be treated fairly in the

---

<sup>1</sup> The parties use the terms “black” and “Hispanic” in their submissions, and the court adopts the parties’ terminology.

<sup>2</sup> See Declaration of Sharon Seeley dated January 21, 2009 (Docket Entry # 253) app. B (citing statistics from the New York City Department of City Planning).

<sup>3</sup> See id.

<sup>4</sup> See id. app. C.

<sup>5</sup> See Declaration of Richard A. Levy dated February 2, 2009 (Docket Entry # 264) Ex. D (showing percentages of minority firefighter representation in 1999 in New York City as 2.9% African-American and 2.8% Hispanic, in Los Angeles as 14.0% African-American and 30.0% Hispanic, in Chicago as 20.4% African-American and 8.6% Hispanic, in Houston as 17.1% African-American and 13.9% Hispanic, and in Philadelphia as 26.3% African-American and 3.2% Hispanic). The 2000 Census figures for those cities show that Los Angeles had 11.2% black residents and 46.5% Hispanic or Latino residents, Chicago had 36.8% black residents and 26.0% Hispanic or Latino residents, Houston had 25.3% black residents and 37.4% Hispanic or Latino residents, and Philadelphia had 43.2% black residents and 8.5% Hispanic or Latino residents. See U.S. Census Bureau, State & County QuickFacts, available at <http://quickfacts.census.gov/qfd/index.html> (last visited on July 21, 2009).

application process for positions in the FDNY. Specifically, the Federal Government and the Intervenor (“Plaintiffs”)<sup>6</sup> challenge the City’s reliance on two written examinations that are used to appoint entry-level firefighters to classes at the New York City Fire Academy (“Academy”). These examinations—Written Examination 7029 and Written Examination 2043—were administered from 1999 to 2007, and the City has appointed more than 5,300 entry-level firefighters based upon their results. Although Plaintiffs identify approximately 3,100 of the examination candidates as black and approximately 4,200 of the examination candidates as Hispanic, the City has appointed just 184 black firefighters and 461 Hispanic firefighters from the challenged examinations. (See Section III.A, *infra*.)

Plaintiffs assert that the City’s reliance on Exams 7029 and 2043 in selecting entry-level firefighters has had a disparate impact on black and Hispanic candidates in violation of Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e, *et seq.* (“Title VII”). The Intervenor also claim, under a disparate treatment theory, that the City, two city agencies, the Mayor and the Fire Commissioner “have long been aware of the discriminatory impact on blacks of their examination process,” and that their “continued reliance on and perpetuation of these racially discriminatory hiring processes constitute intentional race discrimination . . . .” (Intervenor’s Compl. (Docket Entry # 48) ¶ 51.)

To remedy these claimed violations, Plaintiffs seek various forms of injunctive and monetary relief. The Federal Government seeks to enjoin the City from engaging in discriminatory practices “against blacks on the basis of race and against Hispanics on the basis of

---

<sup>6</sup> The Intervenor present claims relating only to black firefighters, while the Federal Government brings claims relating to both black and Hispanic firefighters. Although their Motions are separate, the fundamental issues in each are the same. The court will, therefore, generally refer to both parties as “Plaintiff” throughout, distinguishing between the Federal Government and the Intervenor as needed.

national origin,” and seeks a specific injunction against the practices challenged in this case. (See Compl. (Docket Entry # 1) ¶ 38.) It also asks the court to order the City to take “appropriate action to correct the present effects of its discriminatory policies and practices” and to enjoin it from failing to “make whole” those harmed by the City’s policies and practices. (Id.) The Intervenors seek similar, but broader relief, including an injunction requiring the City to “appoint entry-level firefighters from among qualified black applicants in sufficient numbers to offset the historic pattern and practice of discrimination against blacks in testing and appointment to that position.” (Int. Compl., Prayer For Relief ¶ 3(d).) The Intervenors seek to require the City to “recruit black candidates and implement and improve long-range recruitment programs” and to “provide . . . future test scores, appointment criteria, eligibility lists, appointment data, and all other information necessary to conduct an adverse impact and job-relatedness analysis of the examination and selection process.” (Id. ¶¶ 3(e), (f).) The Intervenors also seek damages and other fees. (Id. ¶¶ 4-9.)

This is not the first time the City has been brought to federal court to defend its entry-level firefighter examinations against charges of discrimination. In the early 1970s, Judge Weinfeld in the Southern District of New York found that the City’s written and physical examinations for entry-level firefighters violated the Equal Protection Clause of the Constitution because of their discriminatory impact on black and Hispanic applicants. See Vulcan Soc’y of New York City Fire Dep’t, Inc. v. Civil Serv. Comm’n, 360 F. Supp. 1265, 1269 (S.D.N.Y. 1973), affirmed in relevant part by 490 F.2d 387 (2d Cir. 1973). Following Judge Weinfeld’s decision in Vulcan Society, the City contracted with a private consulting firm to construct valid written and physical examinations; these contracts were cancelled three years later, however,

apparently on account of a budget crisis. See Berkman v. City of New York, 536 F. Supp. 177, 184 (E.D.N.Y. 1982).<sup>7</sup> At the time of Vulcan Society, Judge Weinfeld noted the “overwhelming disparity between minority representation in the [FDNY] (5%) and in the general population of New York City within the age group eligible for appointment (32%).” Vulcan Soc’y, 360 F. Supp. at 1269. In the three decades that have followed, these minority groups have come to represent an even greater share of the City’s population. Despite these changes, the overwhelmingly monochromatic composition of the FDNY has stubbornly persisted.<sup>8</sup>

This court has already issued several decisions in the case. I have bifurcated the liability and relief phases (see Docket Entry # 47), permitted intervention by the Intervenors (see id.), denied the Intervenors’ motion to amend the Intervenors’ Complaint (see Docket Entry # 182), declined to dismiss the Intervenors’ Complaint on timeliness grounds (see Docket Entry # 231), and certified a class consisting of black applicants for the position of entry-level firefighter (see Docket Entry # 281).

Now before the court are Motions for Summary Judgment by the Federal Government and the Intervenors. (Docket Entries ## 251, 260.) The Federal Government and the Intervenors have moved for summary judgment on the prima facie case of disparate impact, and the Federal

---

<sup>7</sup> In Berkman, female candidates for the job of entry-level firefighter challenged the physical examination component of a prior FDNY test. See Berkman, 536 F. Supp. at 205-16.

<sup>8</sup> By contrast, it has been reported that the composition of the New York City Police Department (“NYPD”) has undergone significant diversification since the 1970s. In 1978, the NYPD was made up of 8.9% black officers and 3.8% Hispanic officers. See Guardians Ass’n of New York City Police Dep’t, Inc. v. Civil Serv. Comm’n, 484 F. Supp. 785, 793 (S.D.N.Y. 1980). According to a New York Times article published in June 2009, the NYPD reported its rank-and-file officers as 18% black and 28.7% Hispanic. See Al Baker, “Police Commissioner Plans to Put More Minority Officers in Top Posts,” N.Y. Times A20 (June 26, 2009), available at <http://www.nytimes.com/2009/06/26/nyregion/26nypd.html> (last visited on July 21, 2009). On July 15, 2009, the NYPD announced that it had sworn in its most diverse Police Academy class ever, including 14.7% black recruits and 33.3% Hispanic recruits. See New York City Police Dep’t, Press Release # 024, dated July 15, 2009, available at [http://www.nyc.gov/html/nypd/html/pr/pr\\_2009\\_024.shtml](http://www.nyc.gov/html/nypd/html/pr/pr_2009_024.shtml) (last visited on July 21, 2009).

Government has joined the Intervenor's Motion for Summary Judgment on the City's business necessity defense.<sup>9</sup> (See Docket Entries ## 251, 260, 263.)

Upon consideration of the parties' submissions and oral argument, the court concludes that Plaintiffs have established a prima facie case that the City's use of the two written examinations has resulted in a disparate impact upon black and Hispanic applicants for the position of entry-level firefighter. The court also concludes that the City has failed to present sufficient evidence supporting a business justification for its employment practices. I therefore grant Plaintiffs' Motions for Summary Judgment in their entirety.

In essence, my ruling is premised upon two basic conclusions. First, Plaintiffs have shown that there is no triable issue of fact as to whether the City's use of Written Exams 7029 and 2043 has resulted in a statistically and practically significant adverse impact on black and Hispanic firefighter applicants. Black and Hispanic applicants disproportionately failed the

---

<sup>9</sup> In support of its Motion, the Federal Government has submitted a Statement of Undisputed Facts Pursuant to Rule 56.1 (Docket Entry # 252) ("USA 56.1"), a Declaration of Sharon A. Seeley (Docket Entry # 253) ("Seeley Decl."), an Affidavit of Bernard R. Siskin, Ph.D. (Docket Entry # 254) ("Siskin Aff."), including a copy of his November 2007 expert report at Exhibit A ("Siskin Report"), a Memorandum of Law (Docket Entry # 255) ("USA Mem."), and a Reply Memorandum (Docket Entry # 258) ("USA Reply"). In response, the City has submitted an opposing Memorandum of Law on the prima facie case (Docket Entry # 256) ("Def. PF Mem.") and a Response to the Federal Government's Rule 56.1 Statement (Docket Entry # 257) ("Def. USA 56.1").

In support of their Motion, the Intervenor's have submitted a Rule 56.1 Statement of Undisputed Facts (Docket Entry # 261) ("Int. 56.1"), a Memorandum of Law (Docket Entry # 262) ("Int. Mem."), a Declaration of Richard A. Levy (Docket Entry # 264) ("Levy Decl."), a Reply Memorandum on the prima facie case (Docket Entry # 268) ("Int. PF Reply") and a Reply Memorandum on job-relatedness and business necessity (Docket Entry # 269) ("Int. BN Reply"). Included in the Levy Declaration are excerpts from the expert report of Dr. Wiesen (Levy Decl. Exs. R, Z ("Wiesen Report").) A full version of the Wiesen Report appears as Docket Entry # 123. In response, the City has submitted an opposing Memorandum of Law relating to the prima facie case ("Def. PF Mem."), an opposing Memorandum of Law relating to job-relatedness and business necessity (Docket Entry # 266) ("Def. BN Mem."), a Response to Intervenor's Rule 56.1 Statement (Docket Entry # 265) ("Def. Int. 56.1"), and a Declaration of William S.J. Fraenkel (Docket Entry # 267) ("Fraenkel Decl."). Included in the Fraenkel Declaration is the expert report of Dr. Schemmer and Dr. Bobko. (Fraenkel Decl. Ex. 1 ("Bobko-Schemmer Report").) The Federal Government has submitted a memorandum joining the Intervenor's on job-relatedness and business necessity (Docket Entry # 263 ("USA BN Mem."), and the City has filed a response (Docket Entry # 272 ("Def. USA Resp.")).

The Intervenor's have also moved to strike two declarations submitted by the City with their summary judgment papers. (See Docket Entries ## 273, 274, 276, 277.) The court addresses this Motion in Section IV.B below.

written examinations, and those who passed were placed disproportionately lower down than white candidates on the hierarchical hiring lists resulting from their scores. Second, although the City has had the opportunity to justify this adverse impact by showing that it used the written examinations to test for the relevant skills and abilities of entry-level firefighters, the City has failed to raise a triable issue on this defense. Under Second Circuit precedent, the evidence presented by the City is insufficient as a matter of law to justify its reliance on the challenged examinations.

Before proceeding to the legal analysis, I offer a brief word about the Supreme Court's recent decision in Ricci v. DeStefano, 129 S. Ct. 2658 (June 29, 2009). I reference Ricci not because the Supreme Court's ruling controls the outcome in this case; to the contrary, I mention Ricci precisely to point out that it does not. In Ricci, the City of New Haven had set aside the results of a promotional examination, and the Supreme Court confronted the narrow issue of whether New Haven could defend a violation of Title VII's disparate treatment provision by asserting that its challenged employment action was an attempt to comply with Title VII's disparate impact provision. The Court held that such a defense is only available when "the employer can demonstrate a strong basis in evidence that, had it not taken the action, it would have been liable under the disparate-impact statute." Id. at 2664. In contrast, this case presents the entirely separate question of whether Plaintiffs have shown that the City's use of Exams 7029 and 2043 has actually had a disparate impact upon black and Hispanic applicants for positions as entry-level firefighters. Ricci did not confront that issue.

The Ricci Court concluded that New Haven would not likely have been liable under a disparate impact theory. See id. at 2681. In doing so, the Court relied on the various steps that

New Haven took to validate its civil service examination. Id. at 2678-79. It is noteworthy, however, that in this case New York City has taken significantly fewer steps than New Haven took in validating its examination. The relevant teaching of Ricci, in this regard, is that the process of designing employment examinations is complex, requiring consultation with experts and careful consideration of accepted testing standards. As discussed below, these requirements are reflected in federal regulations and existing Second Circuit precedent. This legal authority sets forth a simple principle: municipalities must take adequate measures to ensure that their civil service examinations reliably test the relevant knowledge, skills and abilities that will determine which applicants will best perform their specific public duties.

In rendering this decision, I am aware that the use of multiple-choice examinations is typically intended to apply objective standards to employment decisions. Similarly, I recognize that it is natural to assume that the best performers on an employment test must be the best people for the job. But, the significance of these principles is undermined when an examination is not fair. As Congress recognized in enacting Title VII, when an employment test is not adequately related to the job for which it tests—and when the test adversely affects minority groups—we may not fall back on the notion that better test takers make better employees. The City asks the court to do just that. Regrettably, though, the City did not take sufficient measures to ensure that better performers on its examinations would actually be better firefighters. Accordingly, the court grants the Motions for Summary Judgment and finds that Plaintiffs have established disparate impact liability.

## II. SUMMARY JUDGMENT STANDARD

“Summary judgment is appropriate when the pleadings and admissible evidence proffered to the district court show that there is ‘no genuine issue as to any material fact and that the moving party is entitled to a judgment as a matter of law . . . .’” Major League Baseball Props., Inc. v. Salvino, Inc., 542 F.3d 290, 309 (2d Cir. 2008) (quoting Fed. R. Civ. P. 56(c)). “Material facts are those which ‘might affect the outcome of the suit under the governing law,’ and a dispute is ‘genuine’ if ‘the evidence is such that a reasonable jury could return a verdict for the nonmoving party.’” Coppola v. Bear Stearns & Co., 499 F.3d 144, 148 (2d Cir. 2007) (quoting Anderson v. Liberty Lobby, Inc., 477 U.S. 242, 248 (1986)). Factual disputes that are irrelevant or immaterial to the disposition of a case cannot preclude a grant of summary judgment. See Loria v. Gorman, 306 F.3d 1271, 1282-83 (2d Cir. 2002).

In considering a motion for summary judgment, the court construes the facts “in the light most favorable to the nonmoving party,” and draws “all reasonable inferences in its favor.” SCR Joint Venture L.P. v. Warshawsky, 559 F.3d 133, 137 (2d Cir. 2009). “[T]he moving party bears the burden of demonstrating the absence of a genuine issue of material fact.” Baisch v. Gallina, 346 F.3d 366, 371-72 (2d Cir. 2003) (citing Celotex Corp. v. Catrett, 477 U.S. 317, 323 (1986)). In response, the nonmoving party “‘must do more than simply show that there is some metaphysical doubt as to the material facts . . . .’” Jeffreys v. City of New York, 426 F.3d 549, 554 (2d Cir. 2005) (quoting Matsushita Elec. Indus. Co. v. Zenith Radio Corp., 475 U.S. 574, 586 (1986)).

### III. THE PRIMA FACIE CASE

Plaintiffs seek summary judgment on their prima facie case of disparate impact. As set forth below, summary judgment on Plaintiffs' prima facie case is warranted. The facts set forth below are undisputed, unless otherwise noted.

#### A. The Hiring Process

During the relevant period, the hiring of entry-level firefighters from Exams 7029 and 2043 proceeded in several stages. Candidates interested in a position as an entry-level firefighter began by submitting an application to the Department of Citywide Administrative Services ("DCAS"), paid an application fee (unless it had been waived), and received an admission card for a written examination. (See USA 56.1 ¶¶ 13, 14; Int. 56.1 ¶ 9.) Each written examination was an 85-question, paper-and-pencil multiple choice test, and an applicant's raw score on that examination was simply a percentage of the questions answered correctly. (See USA 56.1 ¶¶ 16, 23; Int. 56.1 ¶ 11.) A passing score was set for each examination, and after the results were in, the City notified each applicant of his or her score, as well as whether he or she had passed. (See Int. 56.1 ¶ 17.) Versions of each examination (with the same questions, but sometimes different question-orderings) were administered on repeated occasions—Exam 7029 was administered from 1999 through 2002, and Exam 2043 was administered from 2002 through 2007. (See USA 56.1 ¶¶ 17-22.)

Candidates who passed the written examination were allowed to take the physical performance test ("PPT"), but those who failed the written examination could not take the PPT. (See id. ¶ 27; Int. 56.1 ¶ 14.) The PPT consisted of eight physical tasks, and a candidate had to pass a minimum of six tasks to achieve a passing score overall. (Int. 56.1 ¶ 19.) A passing

candidate's score on the PPT was simply a percentage of the number of tasks successfully completed. For example, passing eight tasks resulted in a score of 100%, passing seven tasks resulted in a score of 87.5%, and passing six tasks resulted in a score of 75%. (See id.)

Candidates who passed both the written examination and the PPT were placed on a "rank-order" eligibility list. (USA 56.1 ¶¶ 28-30; Seeley Decl. app. I.) The ordering of the eligibility list was based upon an elaborate process of, inter alia, "standardizing," "combining," and "transforming" the raw scores. (See USA 56.1 ¶ 31; Int. 56.1 ¶ 20.) Specifically, the raw score from the written examination and the PPT would be "standardized" by subtracting the average score for all candidates from an individual candidate's score and then dividing that number by the standard deviation for the test. (See Seeley Decl. apps. J, K.) The resulting scores from both the written examination and the PPT would then be divided in half and added together to create a "Combined Weighted Standard Score." (Id.) The Combined Weighted Standard Score was then converted into a "Transformed Score" by multiplying by either 18.472906403940886699 (for Exam 7029) or 12.7226 (for Exam 2043), and then adding either 83.74384236453 (for Exam 7029) or 88.4606 (for Exam 2043). (Id.) Finally, the "Adjusted Final Average," used to rank candidates on the eligibility list, was created by adding any "Residency," "Legacy," or "Veteran" points to the Transformed Score. (Id.; see USA 56.1 ¶¶ 30, 31; Int. 56.1 ¶ 20.) This elaborate process resulted in a list of candidates eligible to be appointed to Academy classes in order of rank.

The DCAS and FDNY would determine how many candidates would be needed to fill an upcoming class and would certify a portion of the eligibility list for appointment, beginning with the highest scores. (USA 56.1 ¶¶ 40-41; Int. 56.1 ¶ 23.) The FDNY's Candidate Investigation

Division (“CID”) took steps to process and investigate candidates in order of their ranking on the eligibility list. (USA 56.1 ¶¶ 36-37, 39; Int. 56.1 ¶ 24.) This investigation involved, inter alia, background checks, intake interviews, and medical and psychological examinations by the FDNY’s Bureau of Health Services. (USA 56.1 ¶¶ 35-37; Int. 56.1 ¶ 24; Seeley Decl. app. A (Request for Admission # 101).) The City would fill slots in an Academy class by proceeding down the list of eligible and qualified applicants until the class was filled; once a class was filled, any eligible and qualified candidate still remaining would not be appointed for that class. (USA 56.1 ¶¶ 42, 43; Int. 56.1 ¶ 25.)

Written Examination 7029 was first administered on February 26, 1999, and versions of it were administered as late as 2002. (USA 56.1 ¶¶ 17-19; Int. 56.1 ¶ 15.) Approximately 1,750 black applicants and approximately 2,125 Hispanic applicants sat for Exam 7029. (Siskin Report tbls. 1, 2.) The City hired from the eligibility list resulting from Exam 7029 from February 2001 through at least September 2004, and appointed over 3,200 entry-level firefighters from that examination. (USA 56.1 ¶¶ 11, 44.) Of this number, 104 (3.2%) individuals were black and 274 (8.5%) individuals were Hispanic. (Id. ¶ 11.)

Written Examination 2043 was first administered on December 14, 2002, and versions of it were administered as late as March 2007. (Id. ¶¶ 20-22; Int. 56.1 ¶ 16.) Approximately 1,390 black applicants and approximately 2,125 Hispanic applicants sat for Exam 2043. (Siskin Report tbls. 5, 6.) The City hired firefighters from the eligibility list resulting from Exam 2043 from May 2004 through at least January 2008, and, as of November 2007, the City had appointed over 2,100 entry-level firefighters from that examination. (USA 56.1 ¶¶ 12, 46.) Of this number, 80 (3.7%) were black and 187 (8.7%) were Hispanic. (Id. ¶ 12.)

With these undisputed background facts in mind, the court addresses the prima facie case.

**B. The Use of Statistics for a Prima Facie Case**

A prima facie showing of disparate impact “requires plaintiffs to establish by a preponderance of the evidence that the employer ‘uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin.’” Robinson v. Metro-North Commuter R.R. Co., 267 F.3d 147, 160 (2d Cir. 2001) (quoting 42 U.S.C. § 2000e-2(k)(1)(A)(i)). “To make this showing, a plaintiff must (1) identify a policy or practice, (2) demonstrate that a disparity exists, and (3) establish a causal relationship between the two.” Id. at 160.

Statistics alone can make out the prima facie case. See EEOC v. Joint Apprenticeship Comm. of Joint Indus. Bd. of Elec. Indus., 186 F.3d 110, 117 (2d Cir. 1999); see also Robinson, 267 F.3d at 160 (“[S]tatistical proof almost always occupies center stage in a prima facie showing of a disparate impact claim.”). In order to do so, “[t]he statistics must reveal that the disparity is substantial or significant.” Robinson, 267 F.3d at 160 (internal quotation marks omitted). “Moreover, the statistics must be of a kind and degree sufficient to reveal a causal relationship between the challenged practice and the disparity.” Id. “[A] plaintiff may establish a prima facie case of disparate impact discrimination by proffering statistical evidence which reveals a disparity substantial enough to raise an inference of causation. That is, a plaintiff’s statistical evidence must reflect a disparity so great that it cannot be accounted for by chance.” Joint Apprenticeship Comm., 186 F.3d at 117.

There are at least two widely recognized statistical measures of disparate impact: (1) the 80% or Four-fifths Rule, and (2) statistical significance or standard deviation analysis. See

Atkins v. Westchester County Dep't of Soc. Serv., 31 Fed. Appx. 52, 53 (2d Cir. 2002) (summary order) (“[i]n evaluating disparate impact claims under Title VII, this Court has primarily relied upon [these] two methods of measuring disparities between groups”). Federal regulations set out the 80% Rule, and courts have recognized it as a “rule of thumb” for statistical analysis of disparate impact. See, e.g., Joint Apprenticeship Comm., 186 F.3d at 118 (“This rule is not binding on courts, and is merely a ‘rule of thumb’ to be considered in appropriate circumstances.”); see also United States v. New York City Bd. of Educ., 487 F. Supp. 2d 220, 224 (E.D.N.Y. 2007). The 80% Rule appears at 29 C.F.R. § 1607.4D, which states:

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 Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms . . . .

Id. Essentially, this means that if the minority group performs less than 80% as well as the highest performing group, disparate impact will generally be inferred.

Courts have also relied upon standard deviation analysis (or statistical significance analysis) in determining whether there has been a disparate impact. “Standard deviation analysis measures the probability that a result is a random deviation from the predicted result—the more standard deviations the lower the probability the result is a random one.” Waisome v. Port Auth. of New York & New Jersey, 948 F.2d 1370, 1376 (2d Cir. 1991); see also Barbara Lindemann & Paul Grossman, Employment Discrimination Law (“Lindemann”) 94 (3d ed. 1996) (“Tests of statistical significance are commonly used in the social sciences to rule out chance as the cause

of observed disparities.”). “Basically, looking at standard deviations indicates how far an obtained result varies from an expected result.” Smith v. Xerox Corp., 196 F.3d 358, 365 (2d Cir. 1999), overruled on unrelated grounds by Meacham v. Knolls Atomic Power Lab., 461 F.3d 134, 141 (2d Cir. 2006). The Second Circuit has “looked to whether the plaintiff can show a statistically significant disparity of two standard deviations” in making a prima facie showing. Id. at 365.

“Although courts have considered both the four-fifths rule and standard deviation calculations in deciding whether a disparity is sufficiently substantial to establish a prima facie case of disparate impact, there is no one test that always answers the question.” Id. at 366. According to either measurement, “the substantiality of a disparity is judged on a case-by-case basis.” Id.

#### **C. The Statistics for Plaintiffs’ Prima Facie Case**

Plaintiffs allege that four employment practices related to the challenged examinations have had an unlawful disparate impact on black and Hispanic candidates for the position of entry-level firefighter. Specifically, they challenge the City’s use of:

- (1) Written Examination 7029 as a pass/fail screening device with a cutoff score of 84.705;
- (2) Rank-order processing and selection of candidates from the Written Examination 7029 eligibility list based on a combination of their scores on Written Examination 7029 and the PPT;
- (3) Written Examination 2043 as a pass/fail screening device with a cutoff score of 70;
- (4) Rank-order processing and selection of candidates from the Written Examination 2043 eligibility list based on a combination of their scores on Written Examination 2043 and the PPT.

(USA Mem. 1; see Int. Mem. 7.) As this court previously noted, “[t]he use of an examination as a ‘pass/fail screening device’ means the use [of] the examination to exclude from appointment those applicants that have failed the examination.” (Memorandum & Order dated May 11, 2009 (Docket Entry # 281), at 3.) “The use of an examination as a part of ‘rank-order processing’ means the use of the examination as a component of the overall score that determines an applicant’s position on a hierarchical hiring list.” (Id.)

According to the Federal Government, the statistical significance analysis performed by its expert, Bernard R. Siskin, Ph.D., establishes a prima facie case of disparate impact. (USA Mem. 2.) Similarly, the Intervenors argue that the statistical significance analyses performed by Dr. Siskin and their expert, Joel P. Wiesen, Ph.D., establish a prima facie case. (Int. Mem. 7.)

The court briefly reviews the statistics that Plaintiffs have presented on each of the challenged practices. The City does not dispute the statistical calculations of Plaintiffs’ experts, but rather, disputes Plaintiffs’ reliance on statistical significance testing because of assumptions underlying that methodology. The court will address these assumptions after setting out the undisputed statistical calculations of Plaintiffs’ experts.

**1. Pass/Fail Use of Exam 7029**

The cutoff passing score for Written Examination 7029 was 84.705%. (USA 56.1 ¶ 24; Int. 56.1 ¶ 13.) Based on that cutoff score, the pass rate of white candidates for Exam 7029 was 89.9%, while the pass rate of black candidates was 60.3%. (USA 56.1 ¶ 83; Int. 56.1 ¶ 28.) In other words, out of 12,915 white test takers, 11,613 received a passing score of at least 84.705, whereas out of 1,749 black test takers, only 1,054 received a passing score. (See Int. 56.1 ¶ 27; Wiesen Report, tbl. 3a.) The pass rate of black candidates was, therefore, 67% of the pass rate of

white candidates. (USA 56.1 ¶ 89.) Both Dr. Siskin's and Dr. Wiesen's standard deviation analysis found that this disparity is equivalent to 33.9 units of standard deviation, meaning that the likelihood it occurred by chance is less than 1 in 4.5 million-billion. (See id. ¶ 85 (citing, inter alia, Siskin Report 3, 21); see also Int. 56.1 ¶ 30 (citing, inter alia, Wiesen Report 18-19).)

The practical effect of this disparity, according to Dr. Siskin, is that 519 black candidates who failed the examination—74.7% of the black applicants who failed—were eliminated from consideration. (See USA 56.1 ¶¶ 86-87 (citing Siskin Report).) Dr. Wiesen estimated that 457 black candidates would have passed the examination but for the effect of this disparity. (See Int. 56.1 ¶ 38 (citing Wiesen Report).) Based on Dr. Siskin's calculation, 114 additional black firefighters would have been appointed absent the disparity. (USA 56.1 ¶ 88 (citing Siskin Report).) This last calculation was based on the assumption that the black applicants who failed Exam 7029 would have passed the PPT at the same rate as other similarly situated passers, and would have met the other qualifications and been appointed at the same rate as other passers.<sup>10</sup> (Siskin Report 16-17.)

The pass rate for Hispanic candidates taking Exam 7029 was 76.7%, compared with a pass rate of 89.9% for white candidates. (USA 56.1 ¶ 92.) Accordingly, the pass rate of Hispanic candidates was 85.3% of the pass rate of white candidates. (Siskin Report, tbl. 2.) Dr. Siskin's standard deviation analysis found that this disparity is equivalent to 17.4 units of standard deviation, meaning that the likelihood it occurred by chance is less than 1 in 4.5 million-billion. (USA 56.1 ¶ 94 (citing, inter alia, Siskin Report 3, 23).)

---

<sup>10</sup> The City accepts as undisputed Dr. Siskin's calculation concerning those who would have been appointed absent the disparity, assuming "use of a test of statistical significance, the assumption of 'shortfall' suggested by Dr. Siskin and inferring no difference in the capability and preparedness of the groups being compared." (Def. USA 56.1 ¶ 88; see also id. ¶¶ 86-88, 95-97, 103-05, 111-13.) As discussed below, to the extent that the City challenges these assumptions, the court concludes that its arguments are without merit and are insufficient to preclude summary judgment.

The practical effect of this deviation, according to Dr. Siskin, is that 282 Hispanic candidates who failed the examination—56.9% of the Hispanic applicants who failed—were eliminated from consideration. (See id. ¶¶ 95-96 (citing Siskin Report).) Based on Dr. Siskin's calculation, 62 additional Hispanic firefighters would have been appointed absent the disparity. (Id. ¶ 97 (citing Siskin Report).) This last calculation was based on the assumption that the Hispanic applicants who failed Exam 7029 would have passed the PPT at the same rate as other similarly situated passers, and would have met the other qualifications and been appointed at the same rate as other passers. (Siskin Report 16-17.)

## **2. Pass/Fail Use of Exam 2043**

The cutoff passing score for Written Examination 2043 was 70%. (USA 56.1 ¶ 25; Int. 56.1 ¶ 13.) Based on this cutoff score, the pass rate of white candidates taking Exam 2043 was 97.2%, while the pass rate of black candidates was 85.4%. (USA 56.1 ¶ 100; see Wiesen Report 42.) In other words, out of 13,877 white test takers, 13,495 received a passing score of at least 70, whereas, out of 1,393 black test takers, 1,190 received a passing score. (See Int. 56.1 ¶ 32; Wiesen Report, tbl. 16a.) The pass rate of black candidates was, therefore, 87.8% of the pass rate of white candidates. (Siskin Report, tbl. 5.) Both Dr. Siskin's and Dr. Wiesen's standard deviation analysis found that this disparity is equivalent to 21.8 units of standard deviation, meaning that the likelihood that it occurred by chance is less than 1 in 4.5 million-billion. (See USA 56.1 ¶ 102 (citing, inter alia, Siskin Report 5, 26); see also Int. 56.1 ¶ 35 (citing, inter alia, Wiesen Report 42-43).)

The practical effect of this deviation, according to Dr. Siskin, is that 165 black candidates who failed the examination—81.3% of the black applicants who failed—were eliminated from

consideration. (See USA 56.1 ¶¶ 103-04 (citing Siskin Report).) Dr. Wiesen estimated that 150 black candidates would have passed the examination absent the disparity. (See Int. 56.1 ¶ 39 (citing Wiesen Report).) Based on Dr. Siskin's calculation, 30 additional black firefighters would have been appointed absent the disparity. (USA 56.1 ¶ 105 (citing Siskin Report).) This last calculation was based on the assumption that the black applicants who failed Exam 2043 would have passed the PPT at the same rate as other similarly situated passers, and would have met the other qualifications and been appointed at the same rate as other passers. (Siskin Report 16-17.)

The pass rate for Hispanic candidates taking Exam 2043 was 92.8%, compared with a pass rate of 97.2% for white candidates. (USA 56.1 ¶ 108.) The pass rate of Hispanic candidates was, therefore, 95.5% of the pass rate of white candidates. (Siskin Report, tbl. 6.) Dr. Siskin's standard deviation analysis found that this disparity is equivalent to 10.5 units of standard deviation, meaning that the likelihood it occurred by chance is less than 1 in 4.5 million-billion. (USA 56.1 ¶ 110 (citing, inter alia, Siskin Report 5, 27).)

The practical effect of this deviation, according to Dr. Siskin, is that 94 Hispanic candidates who failed the examination—61.8% of the Hispanic applicants who failed—were eliminated from consideration. (See id. ¶¶ 111-12 (citing Siskin Report).) Based on Dr. Siskin's calculation, 17 additional Hispanic firefighters would have been appointed absent the disparity. (Id. ¶ 113 (citing Siskin Report).) This last calculation was based on the assumption that the Hispanic applicants who failed Exam 2043 would have passed the PPT at the same rate as other similarly situated passers, and would have met the other qualifications and been appointed at the same rate as other passers. (Siskin Report 16-17.)

### 3. Rank-Ordering Use of Exam 7029

On the eligibility list created from Written Exam 7029 and the PPT, black candidates were grouped disproportionately lower down than white candidates. For example, only 10.1% of black candidates were in the top 20% of all candidates, while 53.8% of black candidates were in the bottom 40%, and 29.2% of black candidates were in the bottom 20%.<sup>11</sup> (USA 56.1 ¶ 120.) While 33% of white candidates had eligibility list numbers at or above 2000, only 21% of black candidates did, and while 20% of white candidates had list numbers at or below 5001, 30% of black candidates did. (Id. ¶¶ 118-19.) According to Dr. Wiesen's calculations, the average ranking of a black candidate was 630 ranking places lower than that of a white candidate, amounting to a disparity of 6.5 units of standard deviation. (See Wiesen Report, tbl. 9b.) According to Dr. Siskin's calculation, the disparity between the placement of black and white candidates on the eligibility list is equivalent to 6.5 units of standard deviation, meaning that the likelihood it occurred by chance is less than 1 in 11 billion. (Siskin Report 24, 25; USA 56.1 ¶ 117.) Dr. Siskin calculated that, on account of this disparity, 68 out of 104 black candidates were delayed in appointment for an aggregate total of approximately 20 years of delayed wages and seniority. (USA 56.1 ¶ 122; Siskin Report, tbl. 3b.)

Similarly, the eligibility list created from the Exam 7029 results placed Hispanic candidates disproportionately lower down than white candidates. Only 14.3% of Hispanic candidates were in the top 20% of all candidates, while 47.8% of Hispanic candidates were in the

---

<sup>11</sup> The document containing these figures is an email from a DCAS employee (the "DCAS Email"), and suggests that this lower grouping was more pronounced for the written examination. It shows that only 8.7% of black applicants were among the highest 24.8% of scores on Written Exam 7029, but that 44.3% of black applicants were among the bottom 19.7% of scores. (See Seeley Decl. app. AC.) That is, black applicants had about a third as many of the highest scores and more than twice as many of the lowest scores as would be expected absent a disparity.

bottom 40%, and 27.3% of Hispanic candidates were in the bottom 20%.<sup>12</sup> (USA 56.1 ¶ 130.) While 33% of white candidates had eligibility list numbers at or above 2000, only 28% of Hispanic candidates did, and while 20% of white candidates had list numbers at or below 5001, 29% of Hispanic candidates did. (Id. ¶¶ 128-29.) According to Dr. Siskin's calculation, the disparity between the placement of Hispanic and white candidates on the eligibility list is equivalent to 4.6 units of standard deviation, meaning that the likelihood it occurred by chance is less than 1 in 204,000. (Id. ¶ 127.) Dr. Siskin calculated that, on account of this disparity, 86 out of 274 Hispanic candidates were delayed in appointment for an aggregate total of approximately 23 years of delayed wages and seniority. (Id. ¶ 132; Siskin Report, tbl. 4b.)

#### 4. Rank-Ordering Use of Exam 2043

On the eligibility list created from Written Exam 2043 and the PPT, black candidates were grouped disproportionately lower down than white candidates. For example, only 11.4% of black candidates were in the top 20% of all candidates, while 56.9% of black candidates were in the bottom 40%, and 46.2% of black candidates were in the bottom 20%.<sup>13</sup> (USA 56.1 ¶ 139.) While 28% of white candidates had eligibility list numbers at or above 2000, only 18% of black candidates did, and while 30% of white candidates had list numbers at or below 5001, 50% of black candidates did. (Id. ¶¶ 137-38.) According to Dr. Wiesen's calculations, the average ranking of a black candidate was 974 ranking places lower than that of a white candidate,

---

<sup>12</sup> The DCAS Email shows that only 13.8% of Hispanic applicants were among the highest 24.8% of scores on Written Exam 7029, but that 36.3% of Hispanic applicants were among the bottom 19.7% of scores. (See Seeley Decl. app. AC.) That is, Hispanic applicants had about half as many of the highest scores and about twice as many of the lowest scores as would be expected absent a disparity.

<sup>13</sup> The DCAS Email shows that only 5.3% of black applicants were among the highest 18.1% of scores on Written Exam 2043, but that 48.9% of black applicants were among the bottom 22.3% of scores. (See Seeley Decl. app. AC.) That is, black applicants had less than a third as many of the highest scores and more than twice as many of the lowest scores as would be expected absent a disparity.

amounting to a disparity of 9.6 units of standard deviation. (See Wiesen Report, tbl. 22b.) According to Dr. Siskin's calculation, the disparity between the placement of black and white candidates on the eligibility list was equivalent to 9.5 units of standard deviation, meaning that the likelihood it occurred by chance is less than 1 in 4.5 million-billion. (USA 56.1 ¶ 136.) Dr. Siskin calculated that, on account of this disparity, 44 out of 80 black candidates were delayed in appointment for an aggregate total of approximately fourteen years of delayed wages and seniority. (Id. ¶ 146; Siskin Report, tbl. 12b.)

Similarly, the eligibility list created from the results of Exam 2043 placed Hispanic candidates disproportionately lower down than white candidates. Only 17.2% of Hispanic candidates were in the top 20% of all candidates, while 45.4% of Hispanic candidates were in the bottom 40%, and 24.6% of Hispanic candidates were in the bottom 20%.<sup>14</sup> (USA 56.1 ¶ 152.) While 28% of white candidates had eligibility list numbers at or above 2000, only 25% of Hispanic candidates did, and while 30% of white candidates had list numbers at or below 5001, 39% of Hispanic candidates did. (Id. ¶¶ 150-51.) According to Dr. Siskin's calculation, the disparity between the placement of Hispanic and white candidates on the eligibility list is equivalent to 4.6 units of standard deviation, meaning that the likelihood it occurred by chance is less than 1 in 186,225. (Id. ¶ 149.) Dr. Siskin calculated that, on account of this disparity, 51 out of 187 Hispanic candidates were delayed in appointment for an aggregate total of approximately twelve years of delayed wages and seniority. (Id. ¶ 158; Siskin Report, tbl. 14b.)

---

<sup>14</sup> The DCAS Email shows that only 10.1% of Hispanic applicants were among the highest 18.1% of scores on Written Exam 2043, but that 34.9% of Hispanic applicants were among the bottom 22.3% of scores. (See Seeley Decl. app. AC.) That is, Hispanic applicants had about half as many of the highest scores and about one-and-a-half times as many of the lowest scores as would be expected absent a disparity.

Dr. Siskin also conducted tests addressing the fact that the eligibility list from Exam 2043 was not exhausted, and, therefore, candidates very low down on that list were never reached. Dr. Siskin determined that those who were never reached “effectively failed” the examination process. (See USA 56.1 ¶ 160.) According to Dr. Siskin’s calculations, out of 95 black candidates and 63 Hispanic candidates who would have ranked high enough to be considered for hire, 42 of the black candidates and 28 of the Hispanic candidates would have been appointed absent a disparity resulting from Written Exam 2043. (Siskin Report 33-35; USA 56.1 ¶¶ 145, 157.) Based on the hiring rates of candidates from the Exam 2043 eligibility list, he calculated that the disparity of hiring rates between white and black candidates amounted to 9.7 units of standard deviation, while the disparity of hiring rates between white and Hispanic candidates amounted to 5 units of standard deviation. (USA 56.1 ¶¶ 142, 155.)

Dr. Siskin utilized the data relating those who effectively failed in order to calculate an “effective pass rate” for Written Examination 2043, determined to be 70.3% for white candidates, 41.5% for black candidates, and 58.9% for Hispanic candidates. (Id. ¶¶ 162-63, 167.) This amounts to a statistical disparity between white and black candidates of 21.9 units of standard deviation, and a statistical disparity between white and Hispanic candidates of 10.5 units of standard deviation. (Id. ¶¶ 164, 168.)

#### **D. The Parties’ Respective Positions**

Plaintiffs argue that the presented statistics establish a prima facie case of disparate impact for the four challenged employment practices. (See USA Mem. 10-17; Int. Mem. 6-12.) They point out that the calculated disparities between black and minority candidates resulting from the challenged practices are much greater than three units of standard deviation. They also

emphasize the practical significance of these disparities—for example, the Federal Government relies on the statistical analyses showing that, but for the disparities resulting from the written examinations, “1,060 additional black and Hispanic candidates would have been considered for appointment as FDNY firefighters,” “an estimated 293 additional black and Hispanic candidates would have been appointed as FDNY firefighters,” and “249 black and Hispanic firefighters who were appointed—about 39% of those appointed from the examinations at issue in this case—would have been appointed earlier.” (USA Mem. 2-3.) Finally, Plaintiffs argue that the City has conceded these statistical conclusions. (See, e.g., id. at 3.)

In opposition to the Motions, the City offers several iterations of the same basic argument. In essence, the City asks the court to reject Plaintiffs’ statistical significance analysis because it improperly assumes “perfect parity” among groups of people (see Def. PF Mem. 1-3, 5-7), and erroneously produces a finding of disparate impact solely on account of large sample sizes (see id. at 1, 5, 6, 7). The City asks the court to rely exclusively upon the 80% Rule in determining whether there has been a disparate impact between white and minority candidates. (See id. at 1, 2-3.) Because application of this statistical rule would result in a finding of disparate impact for some, but not all, of the challenged employment practices, the City asks the court to deny summary judgment relating to those practices that do not meet the 80% Rule. (See id. at 7-8.) The City does not contest the specific calculations in Plaintiffs’ Rule 56.1 Statements, instead attacking the assumptions on which they rely, and denying the “materiality” of the facts presented.

Plaintiffs respond that large sample sizes do not undermine the validity of statistical significance testing; rather, they argue, it is small sample sizes that render statistical significance

tests less reliable. (See USA Mem. 18-20; Int. Mem. 10.) Plaintiffs also argue that there is no basis for relying on the 80% Rule to the exclusion of statistical significance testing, and that, in fact, all legal authority is to the contrary. (USA Mem. 20-22; USA Reply 3-7; Int. Mem. 13 n.10; Int. PF Reply 2-4.)

Before addressing the parties' respective positions, the court notes that the dispute regarding the proper statistical measurement for disparate impact does not relate to all of the challenged employment practices. It is undisputed that the City's pass/fail use of Exam 7029 has had a disparate impact upon black candidates under both statistical significance testing and under the 80% Rule. (See USA 56.1 ¶¶ 59, 83, 89.) Moreover, the City offers the 80% Rule only as a means of comparing pass rates, not rank-ordering. (See USA 56.1 ¶¶ 67, 69.) It has not presented an alternative statistical measure for the rank-ordering of candidates. (See Def. PF Mem. 7-8.) The City's preference for the 80% Rule, therefore, solely relates to the pass/fail uses of Exam 2043 with respect to black candidates, and the pass/fail uses of Exam 7029 and 2043 with respect to Hispanic candidates.

**E. Plaintiffs Have Demonstrated a Prima Facie Case**

Plaintiffs have demonstrated a prima facie case of disparate impact by (1) identifying four specific employment practices (each relating to both black and Hispanic applicants), (2) demonstrating that a disparity exists among groups, and (3) establishing a causal relationship between the employment practices and the disparities. See Robinson, 267 F.3d at 160. For each employment practice, Plaintiffs have presented analyses from two experts that thoroughly demonstrate the statistical significance of the disparities between groups of candidates. For each of the pass/fail uses of the examinations, these analyses demonstrate that the disparities between

the pass rates of whites and minority candidates were between 10.5 and 33.9 units of standard deviation. For each of the rank-ordering uses of the examinations, the analyses demonstrate that the disparities between the rankings of whites and minority candidates were between 4.6 and 9.7 units of standard deviation. These statistical disparities show that black and Hispanic candidates disproportionately failed Written Exams 7029 and 2043, and were placed disproportionately lower on the eligibility lists created from those examinations.

The Second Circuit has repeatedly recognized that standard deviations of more than 2 or 3 units can give rise to a prima facie case of disparate impact because of the low likelihood that such disparities have resulted from chance. See Malave v. Potter, 320 F.3d 321, 327 (2d Cir. 2003) (“courts ‘generally consider this level of significance [i.e., two standard deviations] sufficient to warrant an inference of discrimination.’”) (quoting Smith, 196 F.3d at 365); Waisome, 948 F.2d at 1376 (“[a] finding of two or three standard deviations (one in 384 chance the result is random) is generally highly probative of discriminatory treatment”); Ottaviani v. State Univ. of New York, 875 F.2d 365, 372 (2d Cir. 1989) (“It is certainly true that a finding of two to three standard deviations can be highly probative of discriminatory treatment.”); Guardians Assoc. of New York City Police Dep’t, Inc. v. Civil Serv. Comm., 630 F.2d 79, 86 (2d Cir. 1980) (“Guardians”) (“[I]n cases involving large samples, ‘if the difference between the expected value (from a random selection) and the observed number is greater than two or three standard deviations,’ a prima facie case is established.”) (quoting Castaneda v. Partida, 430 U.S. 482, 496 n.17 (1977)). The calculated standard deviations in this case are all well beyond 2 to 3 units, strongly supporting a conclusion of a causal relationship between the observed disparities and the employment practices at issue.

The significance of Plaintiffs' statistics is bolstered by evidence that the disparities have been significant as a practical matter. See Lindemann 94 ("To guard against the possibility that a finding of adverse impact could result from the statistical significance of a trivial disparity or meaningless difference in results, the Uniform Guidelines on Employee Selection Procedures[, 29 C.F.R. § 1607.4D,] and the courts have adopted an additional test for adverse impact: that a statistically significant disparity also has practical significance."). As mentioned above, approximately one thousand additional black and Hispanic candidates would have been considered for appointment as FDNY firefighters had it not been for the disparities resulting from the examinations. Further, absent these disparities, approximately 293 additional black and Hispanic candidates would have been appointed from the eligibility lists used from 2001 through 2008, and approximately 249 black and Hispanic applicants who were actually appointed would have been appointed sooner. Given that, in 2007, the FDNY had 8,998 firefighters, including only 303 black firefighters and 605 Hispanic firefighters (see Seeley Decl. app. C), it is clear that these disparities have a substantial practical significance. In fact, the disparities are overwhelming.

The accuracy of Plaintiffs' statistical calculations is not disputed, and the City's Responses to Plaintiffs' 56.1 Statements essentially concede the statistical picture establishing a prima facie case. The City specifically concedes that: (1) the disparity created by each of the challenged practices is more than three units of standard deviation (see Def USA 56.1 ¶¶ 84, 93, 101, 109, 116, 126, 135, 148; Def. Int. 56.1 ¶¶ 30, 31, 35, 37), (2) the Plaintiffs' calculations of statistical significance are "undisputed" (see, e.g., Def. USA 56.1 ¶¶ 85, 94, 102, 110, 117, 127, 136, 149), and (3) "[o]ne of the City's experts conducted analyses to attempt to verify Dr.

Siskin's statistical calculations and confirmed the results reported by Dr. Siskin" (USA 56.1 ¶ 77; Def. USA 56.1 ¶ 77). Moreover, the City has essentially admitted the calculations performed by Plaintiffs' experts showing the disparities' practical significance.<sup>15</sup> (See Def. USA 56.1 ¶¶ 86-88, 95-97, 103-105, 111-13, 122, 132, 145, 157; Def. Int. 56.1 ¶¶ 37-39; see also Def. USA 56.1 ¶¶ 82, 90, 98, 106 (accepting as undisputed that "a test of statistical significance . . . can result in a finding of disparate impact" for pass/fail uses); id. ¶¶ 114, 123, 134, 147 (accepting as undisputed finding of disparate impact for rank-ordering uses "assuming use of a test of statistical significance").) These admissions eliminate the existence of any factual dispute over the prima facie case.

## **F. The City's Arguments**

### **1. Large Sample Sizes**

Rather than attacking the accuracy of Plaintiffs' statistics, the City objects to Plaintiffs' reliance on statistical significance testing as a general matter. The City raises a number of supposed theoretical problems with such testing. The City's principal argument is that the size of the populations being tested in this case (i.e., the many thousands of applicants who took each examination) renders a statistical significance test unreliable. This is because, the City contends, the "larger the group we examine the more likely we are to find differences[]" among candidates that will cause particular individuals to fail. (Def. PF Mem. 5 (noting that "the larger the group we are examining, the more candidates who sit for the exam, the greater our likelihood that some of them will not do as well as others").)

---

<sup>15</sup> The City's Rule 56.1 Responses regarding practical significance dispute only certain assumptions of Plaintiffs' calculations, but few of these contentions are actually supported by reference to the City's own expert reports or other evidence. The disputes that are offered with such support concern the "shortfall" calculations of Dr. Siskin (see, e.g., Def. USA 56.1 ¶¶ 70-73), and the assumption of "parity" in capability and preparedness among racial and ethnic groups (see, e.g., id. ¶¶ 86-88, 95-97, 103-05, 111-13). The court addresses these assumptions below.

The City has it backwards. Rather than undermining confidence in statistical significance testing, large sample sizes make such testing more reliable. Larger sample sizes create a greater likelihood that random differences between individuals will even out among all groups, and a lower likelihood that significant differences between the performance of racial or ethnic groups will have resulted from chance. Existing precedent confirms this principle. Courts have sometimes declined to rely on statistical significance analysis when a sample size was too small. See, e.g., Lindemann 1734 (“Courts have recognized that statistical evidence often is unreliable when the sample size is small.”); Pietras v. Bd. of Fire Comm’rs, 180 F.3d 468, 475 (2d Cir. 1999) (recognizing “authority holding that a disparate impact finding based solely on a sample size as small as the one presented here [i.e., 7 people] cannot stand”). Yet, the City has pointed to no cases rejecting such testing because a sample size was too large. As the Second Circuit stated in Guardians, “in cases involving large samples, ‘if the difference between the expected value (from a random selection) and the observed number is greater than two or three standard deviations,’ a prima facie case is established.” 630 F.2d at 86 (emphasis added) (quoting Castaneda, 430 U.S. at 496 n.17).

The City’s own admissions support this understanding of large sample sizes. The Federal Government has provided a helpful illustration that the City explicitly accepts as undisputed:

Flipping a coin is a common example that illustrates why sample size should affect the number of standard deviations that is equivalent to a given disparity. Flipping a fair coin 10 times will not always result in exactly five heads and five tails; a result of six heads and four tails on ten flips would not indicate with a reasonable degree of certainty that the coin was not fair (i.e., that the disparity was not likely due to chance variation). However, if one flipped a fair coin 1,000 times, one would expect that the number of heads and tails would be close to equal, and a result of 600 heads and 400 tails would allow one to conclude with a high degree of certainty that the coin was not fair (i.e., that

disparity between the rate at which heads came up and the rate at which tails came up was not likely do to chance variation).

Put simply, with a disparity in pass rates of a given size, the bigger the sample (e.g., the more times one flips the coin, or the more applicants who take the test), the more confident one can be that the difference in pass rates in the sample is not due to chance.

(USA 56.1 ¶¶ 56-57 (internal citations omitted); see Def. USA 56.1 ¶¶ 56-57).)

Another undisputed statement, which relies on the City's own expert, further supports the reliability of statistical significance testing when large sample sizes are involved:

With a large sample size, a test of statistical significance using 1% as the standard (i.e., concluding that there is a statistically significant disparity if there is no more than a 1% likelihood of observing a disparity so large due to chance) is better than the 80% Rule at controlling for false positives (situations in which the test used will indicate a disparity when there is no disparity) and false negatives (situations in which the test will indicate there is no disparity when there is a disparity). In other words, with a large sample size, a test of statistical significance is more likely to produce the "right" answer to the question of whether there is a non-chance disparity between the pass rates of two groups.

(USA 56.1 ¶ 64 (citing deposition of City's expert, Dr. Bobko) (emphases added); see Def. USA 56.1 ¶ 64).) These undisputed statements plainly support a conclusion that large sample sizes enhance, rather than undermine, the reliability of statistical testing. Accordingly, while the City purports to challenge the use of statistical significance testing based on sample size, the City's own admissions contradict its position.<sup>16</sup>

When an employment examination is used to make hiring decisions for thousands of applicants, seemingly small differences in pass rates can have a substantial effect on large groups

---

<sup>16</sup> Moreover, the Federal Government points out that Dr. Siskin has performed a recalculation of his statistical significance testing by reducing the sample sizes by 90%. (USA Mem. 18 n.13.) As the City's Bobko-Schemmer Report concedes, this recalculation based on a greatly reduced sample size does not affect the finding of statistical significance for the pass/fail uses of Exams 7029 and 2043. (See Bobko-Schemmer Report 17-18; see also Siskin Report 22, 23, 26-27, 27-28.) Although this recalculation for some of the rank-ordering disparities results in less than three units of standard deviation, the City provides no alternative statistical measure with respect to rank-ordering, and so the recalculation does not undermine the usefulness of the statistics before the court.

of people. Contrary to the City's position, therefore, it is important to rely upon statistical testing to determine whether such differences have resulted from chance or, rather, from a particular employment practice. In this case, statistical significance testing has been used to show that the disparities between groups of candidates have resulted from the challenged examinations. The City's arguments to the contrary are unavailing.

2. Perfect Parity Among Groups

The City also attacks Dr. Siskin's "shortfall" analysis, which estimates the number of minority candidates who would have passed or been appointed had the written examinations not had a discriminatory impact. The City criticizes the fact that such calculations hypothesize a world of "perfect parity" among racial or ethnic groups. (Def. PF Mem. 5.) In other words, the City argues that Plaintiffs' analyses inappropriately compare the racial disparity in test results to a hypothetical world in which racial and ethnic groups perform equally well. (See *id.* ("[S]tatistical significance testing will assume that all people perform at equal levels. However, we know that all individuals do not perform at the same level.")) This argument misstates the law.

First of all, the court rejects the premise that comparison to a standard of equality among groups provides an improper foundation for statistical testing under Title VII. In order to determine whether a particular employment practice has had a disparate impact on a minority group, statistical tests "ask what the results would be for the salient variable . . . if there [had been] no discrimination." *Adams v. Ameritech Servs., Inc.*, 231 F.3d 414, 424 (7th Cir. 2000) (emphasis added). To determine what results "would be," statistical tests properly assume that racial or ethnic groups will perform equally well absent discrimination. See *Smith*, 196 F.3d at

366 (recognizing “null hypothesis” of no difference between compared groups).<sup>17</sup> Statistical significance testing relies on this assumption of equality in assessing whether disparities among groups are based upon chance, or rather, upon some other factor, such as race or national origin. See Ottaviani, 875 F.2d at 371 (“Statistical significance is a measure of the probability that a disparity is simply due to chance, rather than any other identifiable factor.”); Joint Apprenticeship Comm., 186 F.3d at 117 (“a plaintiff’s statistical evidence must reflect a disparity so great that it cannot be accounted for by chance”); Lindemann 94 (“Tests of statistical significance are commonly used in the social sciences to rule out chance as the cause of observed disparities.”). In accordance with these principles, Plaintiffs’ statistical evidence shows that the disparities in this case have not been the result of chance; instead, the disparate impact upon black and Hispanic candidates has resulted from the challenged employment practices.

The court similarly rejects the suggestion in the City’s Rule 56.1 Responses that a prima facie case has not been established because disparities between white, black, and Hispanic candidates can be explained by differences in their “capability and preparedness.” (See, e.g., Def. USA 56.1 ¶¶ 86-88, 95-97, 103-05, 111-13.) The City’s Rule 56.1 Response suggests that the City believes black and Hispanic candidates received lower scores on its written examinations because of their lower capability and preparedness for the job of firefighter. But, if the City contends that differences in aptitude relating to the job of firefighter have led to an adverse impact on minority groups, Title VII’s burden-shifting framework allows the City to

---

<sup>17</sup> By way of further example: “[i]f the relevant market is 40% African-American, for instance, one would expect 40% of hires to be African-American . . . . If the observed percentage of African-American hires is only 20%, then the statistician will compute the ‘standard deviation’ from the expected norm and indicate how likely it is that race played no part in the decisionmaking. Two standard deviations is normally enough to show that it is extremely unlikely (that is, there is less than a 5% probability) that the disparity is due to chance, giving rise to a reasonable inference that the hiring was not race-neutral; the more standard deviations away [from zero], the less likely the factor in question played no role in the decisionmaking process.” Adams, 231 F.3d at 424.

justify the disparate impact as a matter of business necessity. At the prima facie stage, however, the question is only whether there are disparities attributable to the challenged practices, not whether the City can provide a justification for them. During this stage, the City cannot rebut the existence of disparities by claiming that they are explained by the overall “capability and preparedness” of particular groups.

Therefore, the court rejects the City’s arguments about the assumptions in Plaintiffs’ statistics.

3. The 80% Rule

Finally, in opposing summary judgment, the City argues that the court should rely on the 80% Rule to the exclusion of statistical significance testing. There is no support for this position. Controlling precedent holds that the 80% Rule is not an exclusive means of proof, and that alternative statistical tests should be considered. See Joint Apprenticeship Comm., 186 F.3d at 118 (“[80%] rule is not binding on courts, and is merely a ‘rule of thumb’ to be considered in appropriate circumstances.”); see also Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 995 n.3 (1988) (citing criticism of the 80% Rule, recognizing the usefulness of statistical methods in Title VII cases, and endorsing a case-by-case approach); Bew v. City of Chicago, 252 F.3d 891, 893 (7th Cir. 2001) (“The district court properly noted that the 80% guideline may be ignored when other statistical evidence indicates a disparate impact.”). Moreover, the regulation containing the 80% Rule plainly endorses the use of alternative tests. It specifically states: “Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms. . . .” 29 C.F.R. § 1607.4D (emphasis added). Rather than supporting the use of the 80% Rule to the exclusion of statistical significance testing,

this sentence expressly contemplates alternative statistical tests as a means of showing disparate impact. This is precisely the showing Plaintiffs have made in this case.

In support of its preference for the 80% Rule, the City points to the Second Circuit's decision in Waisome. In Waisome, the plaintiffs challenged examinations used by the Port Authority in the process of promoting police officers. 948 F.2d at 1372. The standard deviation between the pass rates of white and black candidates in that case was 2.68, and the district court concluded that this was insufficient to find disparate impact because the pass rate of blacks was 87.2% the pass rate of whites. Id. at 1375. In approving this part of the district court decision, however, the Second Circuit did not find that the 80% Rule controlled the outcome of the case. Instead, the court also relied upon the fact that the calculated statistical significance, 2.68 standard deviations, was a borderline figure, and that, as a practical matter, "if two additional black candidates passed the written examination the disparity would no longer be of statistical importance." Id. at 1376.

The statistics in this case, however, are uniformly beyond 3 units of standard deviation, and, for many of the analyses performed, drastically beyond that. There are also large sample sizes, which make a finding of statistical significance more reliable. Hundreds of black and Hispanic applicants were affected by the City's written examinations. Given the practical significance of the disparities here on the actual hiring rates for black and Hispanic applicants, this case is clearly distinguishable from Waisome. Finally, it is worth noting that the Second Circuit remanded the case in Waisome because it found that there had been a showing of disparate impact. Id. at 1372.

In sum, the City has conceded the accuracy of the calculations of Plaintiffs' experts, which provide ample support for the statistical and practical significance of the disparities at issue. The City's only defense is to resort to abstract arguments relating to the nature of statistical testing in general. (See, e.g., Mar. 19, 2009 Tr. 29-30 (arguing against statistical significance testing based upon Plato's "allegory of the cave").) At the same time, the City has made admissions in its Responses to Plaintiffs' Rule 56.1 Statements that directly contradict its only purported challenges to Plaintiffs' proof. Considering its crucial admissions and the lack of legal authority for its position, the City has failed to wage a serious attack on Plaintiffs' prima facie case.

Under these circumstances, the court finds no material factual dispute relating to the prima facie case. To the extent the City purports to dispute the factual evidence presented by Plaintiffs, it has raised nothing more than metaphysical doubts about the nature of that evidence. Such doubts cannot preclude summary judgment. See Matsushita, 475 U.S. at 586. There is no dispute that Plaintiffs have satisfied a statistical standard for a prima facie case of disparate impact that has been repeatedly accepted by the Second Circuit, nor is there any dispute that they have shown that the disparity has had a substantial, practical significance for the composition of the eligibility lists and hiring of entry-level firefighters. Accordingly, the court grants summary judgment for Plaintiffs on their prima facie case of disparate impact.

#### **IV. BUSINESS NECESSITY**

While Plaintiffs have shown that the City's uses of Written Examinations 7029 and 2043 resulted in a disparate impact upon black and Hispanic candidates, the City may defend against Title VII liability by showing that those uses were justified by legitimate business and job-related

considerations.<sup>18</sup> The City bears the burden of making this showing.<sup>19</sup> See Gulino v. New York State Educ. Dep't, 460 F.3d 361, 385 (2d Cir. 2006). In Gulino, the Second Circuit explained the business necessity defense as follows:

[T]he basic rule has always been that “discriminatory tests are impermissible unless shown, by professionally acceptable methods, to be predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.” This rule operates as both a limitation and a license for employers: employers have been given explicit permission to use job related tests that have a disparate impact, but those tests must be “demonstrably a reasonable measure of job performance.”

460 F.3d at 383 (quoting Albemarle Paper Co. v. Moody, 422 U.S. 405, 431, 426 (1975)).

In setting forth its analysis, this court first reviews the process by which the City created the two challenged examinations. The court then addresses a motion to strike post-discovery submissions made by the City relating to its business necessity defense. Finally, the court sets

---

<sup>18</sup> “Though the terms ‘business necessity’ and ‘job related’ appear to have semantic differences, they have been used interchangeably by the courts.” Gulino v. New York State Educ. Dep't, 460 F.3d 361, 382 (2d Cir. 2006).

<sup>19</sup> The basic disparate impact burden-shifting framework suggests that the determination of whether there is a sufficient disparity is limited to the prima facie case, and that, once that burden is satisfied, the City must show job-relatedness. Nevertheless, Gulino and Robinson unmistakably state that, following a prima facie showing, a defendant may still “directly attack plaintiff’s statistical proof by pointing out deficiencies in data or fallacies in the analysis.” Gulino, 460 F.3d at 382; see also Robinson, 267 F.3d at 161. At that stage, “[t]o successfully contest the plaintiffs’ statistical evidence . . . the employer has to convince the fact finder that its numerical picture is more accurate, valid, or reliable than the plaintiffs’ evidence.” Robinson, 267 F.3d at 161 (internal quotation marks and alteration omitted). The employer bears the burden of this showing.

The parties do not explicitly address this component of the City’s burden, as their arguments relating to statistics are made entirely in the context of the prima facie case. However, the City does argue that “trial courts have discretion to decide whether to use the 80% Rule of statistical significance testing in deciding Title VII claims,” and that the court “should choose [to] employ the 80% Rule . . . in the instant matter.” (Def. PF Mem. 3, 4.) Thus, the City asks the court to proceed to trial on the issue of whether “its numerical picture is more accurate, valid, or reliable than the plaintiffs’ evidence.” Robinson, 267 F.3d at 161. But, as discussed above, the court need not conduct a trial to resolve the question of whether the 80% Rule should be used to the exclusion of Plaintiffs’ statistical tests, when the controlling legal authority has already answered that question. Moreover, to the extent that the City’s position attempts to show “fallacies in the analysis” of Plaintiffs’ experts, see Gulino, 460 F.3d at 382, its arguments have come entirely in the form of unsupported attacks on the usefulness of statistical significance testing in general. Besides these attacks, discussed and rejected above, the City has completely conceded Plaintiffs’ statistical picture. With no factual dispute on which to conduct a trial, summary judgment is appropriate.

out and applies the relevant Second Circuit standard for assessing the challenged examinations. The court concludes that the City has not met that standard.

**A. Creation of Challenged Examinations**

The City used the same development process for Written Exams 7029 and 2043. For Written Exam 7029, a “Test Development Report” was prepared by Matthew Morrongiello (“Morrongiello”), a Tests and Measurement Specialist in the City’s DCAS. (See Levy Decl. Ex. EE (“Test Development Report”).) Alberto Johnston (“Johnston”) of DCAS was “primarily responsible” for developing Exam 2043, and testified that he “was told that we probably can use the old job analysis . . . from [Exam] 7029 . . . .” (Fraenkel Decl. Ex. 6. (“Johnston Dep.”), at 17-18, 19.) Accordingly, the same Test Development Report was relied upon in the City’s development of Written Exam 2043. (See Int. 56.1 ¶ 12; see also *id.* ¶¶ 85-86; Levy Decl. Ex. CC, at 5 (Admission # 11); Fraenkel Decl. Ex. 10 (“Patitucci Dep.”), at 208-09.)<sup>20</sup>

The Test Development Report is a twelve-page document with nine appendices that sets out the process by which the City arrived at the abilities it intended to evaluate on the written examinations, as well as the number of questions it would devote to each ability. The salient features of the process, as set out in that Report, are not in dispute.<sup>21</sup> The Test Development

---

<sup>20</sup> The City also points to a draft report by Dr. Frank Landy entitled, “Job Analysis and Written Examination Development For the City of New York Firefighter Examination No. 0084” dated December 18, 1992. (See Frankel Decl. Ex. 5 (“Landy Report”); see also Def. BN Mem. 5 (relying on the Landy Report).) The Landy Report was prepared in the course of developing an earlier examination, Written Examination 0084. (See Test Development Report 4.) Morrongiello, who was responsible for developing Exam 7029 and the Test Development Report, stated in his deposition that he relied on the Landy Report “to a degree.” (See Fraenkel Decl. Ex. 22 (“Morrongiello Dep.”), at 478-79.) Specifically, he stated that he used the task list in that report as a starting point from which to begin developing Exam 7029. (See *id.* at 478.) As indicated below, the Test Development Report supports this assertion.

<sup>21</sup> The Intervenors’ Rule 56.1 Statement sets out the various steps in the process described in the Test Development Report, relying largely on the Report but supplementing its statements with other support in the record. (See, e.g., Int. 56.1 ¶¶ 87-90.) In response, the City has denied most of the assertions and cited the Test Development Report

Report states that a series of meetings were held, pursuant to which DCAS concluded that it would conduct a new, “comprehensive” job analysis, determine how many people would be needed for a “job analysis survey,” and convene a panel of 8 to 12 incumbent firefighters to review the task and ability lists and the job analysis questionnaire.<sup>22</sup> (See Test Development Report 4.) The basic plan for the process was as follows: first, the tasks and abilities most relevant to the job of firefighter would be determined; second, the relative importance of these tasks and abilities would be assessed; third, clusters of tasks would be matched up with the abilities needed to perform them; and finally, a test would be created to evaluate the identified abilities in the proper proportions. The court briefly reviews this process, as set forth in the Test Development Report.

1. Deriving a List of Tasks and Abilities

In order to familiarize himself with the job of firefighter, Morrongiello conducted interviews with six incumbent firefighters. (Id. at 5.) He used the results of these interviews, coupled with the task list used to develop a prior examination, Exam 0084, to create an “updated task list” reflecting the tasks performed by firefighters. (Id.) Morrongiello then compiled a list of 21 cognitive abilities, derived from “Fleishman’s ability list.” (See Test Development Report 5; Levy Decl. Ex. X, at 129.) Morrongiello convened a focus group of ten firefighters who reviewed the task and ability lists, and, based on the focus group, “suggested changes to the

---

“for a full and accurate statement of its contents and the steps taken by [] Morrongiello.” (See, e.g., Def. Int. 56.1 ¶¶ 87-90.) There is no dispute about the process, but only about the significance of the steps taken.

<sup>22</sup> The written examinations were not intended to include any measure of physical ability, because these would be tested on a separate physical examination, presumably the PPT. (Test Development Report 4.) The Test Development Report states an intention to include questions on the job analysis questionnaire about Oral Comprehension, Written Comprehension, Oral Expression, and Written Expression. (Id.)

proposed task list so it more accurately reflected the current job of Firefighter.” (Test Development Report 5.)

The focus group also reviewed a Job Analysis Questionnaire (“JAQ”) that was distributed to 195 incumbent firefighters. (See id. at 5, 6, 10.) The JAQ asked the firefighters to assess whether 196 listed tasks (grouped into 21 specific task clusters, plus a miscellaneous cluster) were “4. Critical,” “3. Important,” “2. Somewhat important,” or “1. Not [] important” to the job of firefighter, and, similarly, whether 21 listed abilities were “4. Critical,” “3. Important,” “2. Somewhat important,” or “1. Not relevant” to the firefighter job. (Id. at 10 & app. E (JAQ test results).) The responses to the JAQ by 192 of the 195 surveyed firefighters, excluding 3 defective responses, were used to hone the list of abilities down to 18, and the list of tasks down to 111. (Id. at 10.) This was done by removing those tasks and abilities which did not receive at least an average score of 2.5, corresponding to a rating between “important” and “somewhat important.” (Id.)

## 2. Linking Tasks With Abilities

With these results in hand, Morrongiello assembled twelve firefighters into a “Linking Panel,” whose purpose was to “rate—or link—the task clusters to the abilities.” (Id.) The specific “clusters” of tasks important to the job of entry-level firefighter were:

- Initial Response to Incidents/Driving (tasks that “occur between receiving an alarm and initial fire fighting or emergency activities, including driving apparatus to and from various points”);
- Size Up (tasks that involve “evaluating the fire or incident scene to determine actions which should initially be taken and obtaining information needed for evaluation”);
- Ladder Operation (tasks that involve “stabilizing ladder trucks and elevating and operating aerial ladders and platforms in order to rescue victims, provide access for ventilation, operate master stream devices, etc.”);

- Climbing and Portable Ladder Activities (tasks that involve “climbing ladders, stairs and fire escapes, and raising and setting up portable ladders”);
- Building Entry (tasks that involve “prying open or breaking through doors or otherwise entering buildings in order to search for and rescue victims and provide access to the fire for offensive fire fighting, using axes, halligan tool, hooks, rabbit tools, sledge hammers, power saws, and other tools”);
- Search (tasks that involve “searching fire or assigned area in order to locate victims and to obtain further information about fire, following standard search procedures”);
- Rescue (tasks that involve “assisting, carrying or dragging victims from emergency area by means of interior access (stairs, hallways, etc.) or, if necessary, by ladders, fire escapes, platforms, or other means of escape”);
- Ventilation (tasks that involve “opening or breaking open windows, chopping or cutting holes in roofs, breaking through walls or doors, and hanging fans in windows or doors to remove heat, smoke and gas from burning buildings”);
- Supplies Water for Hose Operation (tasks that involve “connecting or hooking up engine to fire hydrant and operating pumps to supply water in appropriate pressure and volume for fire fighting, using hydrant wrenches, couplings, hoses, spanner wrenches, and other tools”);
- Hose Operations During Extinguishment (tasks that involve “stretching line to fire scenes and delivering water to scene of fire”);
- Overhaul (tasks that involve “opening up walls and ceilings, cutting or pulling up floors and moving or turning over debris, in order to check for hidden fires which could rekindle or spread, using hooks, axes, saws and pitchforks”);
- Salvage (tasks that involve “moving and covering furniture, appliances, merchandise and other property, and covering holes in buildings and redirecting or cleaning up water in order to minimize damage, using plastic and canvas covers, ropes, staple guns, mops, squeegees, and other tools”);
- Clean Up/Pick Up (tasks that involve “picking up and returning equipment to vehicle and rolling up or folding up hose, so that the company can go back in service”);
- Equipment Maintenance (tasks that involve “inspecting, cleaning, and maintaining apparatus, equipment carried on the apparatus, and personal gear and equipment”);
- Inspection of Buildings/Hydrants (tasks that involve “inspecting buildings for code violations or hazards on a periodic basis or during the course of activities, and inspecting hydrants for operational use”);
- Extrication (tasks that involve “extricating victims from vehicles, cave-ins, collapsed buildings or other entrapments in order to save lives, using shovels, torches, drills, pry bars, saws, jacks, hurst tools, air bags, and other equipment”);

- Providing Medical Assistance (tasks that involve “providing first aid and direct medical assistance to persons requiring emergency attention”);
- Elevator Related Tasks (tasks that involve “controlling elevators and rescuing persons from stalled elevator cars”);
- Training (tasks that involve “participating in drills which simulate important fire or rescue activities, and attending lectures or formal training”);
- Watch Duties (tasks that involve “standing watch to receive incoming alarms and information, answering phones, and monitoring access to the station house”);
- Station Duties and Chores (tasks that involve “performing routine housekeeping chores or ‘committee work’”); and
- Miscellaneous (tasks that involve “miscellaneous tasks”).

(See id. app. E, at USA000371 – USA000380; see also Levy Decl. Ex. FF (“Linking Panel Worksheet”).) The 18 abilities to which members of the Linking Panel were supposed to “link” these tasks were:

- Oral Comprehension (the ability to “understand spoken English words and sentences”);
- Written Comprehension (the ability to “understand written sentences and paragraphs”);
- Oral Expression (the ability to “use English words or sentences in speaking so that others will understand”);
- Written Expression (the ability to “use English words or sentences in writing so that others will understand”);
- Fluency of Ideas (the ability to “produce a number of ideas about a given topic”);
- Originality (the ability to “produce unusual or clever ideas about a given topic or situation,” and to “invent creative solutions to problems or to develop new procedures for situations in which standard operating procedures do not apply”);
- Memorization (the ability to “remember information, such as words, numbers, pictures and procedures. Pieces of information can be remembered by themselves or with other pieces of information”);
- Problem Sensitivity (the ability to “tell when something is wrong or is likely to go wrong. It includes being able to identify the whole problem as well as elements of the problem”);
- Deductive Reasoning (the ability to “apply general rules to specific problems to come up with logical answers. It involves deciding if an answer makes sense”);

