UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK

----------X

UNITED STATES OF AMERICA,

Plaintiff,

-and-

THE VULCAN SOCIETY, INC., ET AL.,

Plaintiffs-Intervenors,

-against-

THE CITY OF NEW YORK, ET AL.,

Defendants.

----------X

Case No. CV 07 2067 (NGG) (RLM)

ECF Case

MEMORANDUM OF LAW IN SUPPORT OF PLAINTIFFS-INTERVENORS' MOTION FOR SUMMARY JUDGMENT

CENTER FOR CONSTITUTIONAL RIGHTS
666 Broadway, 7th Floor
New York, NY 10012-2399
(212) 614-6438
(212) 614-6499 (fax)

LEVY RATNER, P.C.
80 Eighth Avenue, 8th Floor
New York, NY 10011
(212) 627-8100
(212) 627-8182 (fax)

SCOTT + SCOTT, LLP
29 West 57th Street
New York, NY 10019
(212) 223-6444
(212) 223-6334 (fax)

Attorneys for Plaintiffs-Intervenors

On the brief:
Richard A. Levy
Dana Lossia
Robert H. Stroup
Darius Charney

Date of Service: February 2, 2009

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page

Federal Cases

Federal Statute

Federal Rules

Federal Regulations

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK
----------X
UNITED STATES OF AMERICA,

Plaintiff,

-and-

THE VULCAN SOCIETY, INC., ET AL.,

Plaintiffs-Intervenors,

-against-

THE CITY OF NEW YORK, ET AL.,

Defendants.
----------X

MEMORANDUM OF LAW IN SUPPORT OF PLAINTIFFS-INTERVENORS' MOTION FOR SUMMARY JUDGMENT

Case No. CV 07 2067 (NGG) (RLM)

ECF Case

INTRODUCTION AND SUMMARY OF ARGUMENT

Plaintiffs-Intervenors submit this memorandum of law in support of their motion for summary judgment. This Title VII case challenges the unlawful adverse impact upon black applicants of the cognitive only written exams used by the City of New York to screen and rank applicants for entry-level firefighter positions. The challenged exams – Exams 7029 and 2043 – were first given in 1999 and 2002 respectively.

By way of background, there has been, for decades, a dramatic disproportion between the three (3%) percent of black firefighters in the FDNY and the roughly twenty-seven (27%) percent black population of New York City. The challenged cognitive exams (and similar predecessor exams) have been a major factor in excluding black citizens from firefighter positions in the City of New York. The City's own witnesses have verified that the FDNY has the worst record of diversity of any New York City agency, and when compared to other fire departments in major U.S. cities.

Experts in the field of industrial/organizational psychology and test development have long known that cognitive abilities, above a threshold level, are not, by and large, the abilities of greatest importance in the performance of a firefighter's job. And they have long known that tests of those abilities have the greatest adverse impact on minorities. *Non-cognitive* attributes, such as teamwork, integrity, resilience, tolerance for stress and adaptability, are generally as good as or better than most cognitive abilities in predicting firefighter performance and have less adverse impact, and valid measures of such attributes have been available for decades.

This case can easily be decided on summary judgment because Defendants' experts and employees have either conceded or simply failed to contest every material fact relevant to such a decision. The legal principles governing this challenge to Defendants' selection procedures are well-established. A plaintiff's initial burden is to demonstrate, usually through statistical evidence, a significant enough adverse impact upon minority applicants to establish a *prima facie* case of discrimination. *Gulino v. New York State Bd. of Educ.*, 460 F.3d 361, 382 (2d Cir. 2006). The burden then shifts to the defendant to show that the challenged exam is "validated," i.e., that it is "job related" and consistent with "business necessity." Id. If the defendant succeeds, a plaintiff may still prevail by demonstrating that available alternative measures would have been equally valid but with less adverse impact. Id.

ADVERSE IMPACT:

Plaintiffs-Intervenors' statistical proof that Written Exams 7029 and 2043 had an adverse impact on black applicants is conceded. The City admits that, using generally accepted analyses, the statistical disparity between blacks and whites, in terms of both pass rates and ranking, is greater than three (3) units of standard deviations on both Exams 7029 and 2043. As this is the standard accepted by this Circuit, Plaintiffs-Intervenors' *prima facie* case is established.

VALIDITY:

To meet their burden of proving that the challenged exams were "job related" and consistent with "business necessity," Defendants submitted a three (3) page report that glaringly omitted any opinion that the tests were valid or could be validated as used. This was not an oversight. Defendants' expert, Dr. Philip Bobko, candidly admitted at deposition that his expert report did *not* establish the validity of Defendants' use of Exams 7029 and 2043 as pass/fail screening devices, with cutoff scores of 84.705 and 70.000, respectively, or for ranking candidates. Thus, it is not possible for Defendants to carry their burden of proving "job relatedness" and "business necessity."

The expert report submitted by Plaintiffs-Intervenors, and the three (3) expert reports submitted by the United States, show that Exams 7029 and 2043 were fatally flawed in their construction and use because, *inter alia*, (a) the pas/fail cutoff scores and the rank-ordering of candidates were not shown to predict performance; (b) the written tests did not measure the most important abilities and attributes needed by firefighters; (c) the test construction did not meet professional standards; and (d) the reading level of the exams was too high, obscuring the abilities purportedly being measured. Defendants did not respond to these expert findings. Since they cannot introduce new evidence now (three months after fact and expert discovery has closed), there is simply no disputed fact or opinion which would deter summary judgment.

AVAILABLE ALTERNATIVES:

Even if Defendants had made some showing of validity, summary judgment would still be required since, at the time the challenged exams were given, in 1999 and 2002, there were *numerous* alternatives available, which would have had equal or greater validity and less adverse impact on black candidates. The availability of alternative (non-cognitive) measures to predict

firefighter performance was made widely known by the U.S. Civil Service Commission in the late 1970s, and such measures have been used ever since. They include tests for such attributes as responsibility, desire to learn, teamwork, getting along with people, and many others. Significantly for this motion, it is *Defendants'* own employees, experts and documents that confirm that these alternative measures were available years before Exams 7029 and 2043 were administered.

In fact, after this litigation began, Defendants administered Exam 6019 which tested for a host of non-cognitive personal attributes using "situational judgment exercises" in a written, multiple choice exam. Defendants' witnesses conceded that Exam 6019 had equal or greater validity than the challenged exams and less adverse impact on black applicants *and* that the non-cognitive attributes it measured could have been tested on written exams in 1999 and 2002.

In short, there are no material facts in dispute and the challenged exams violated Title VII as a matter of well-settled law. We respectfully refer the Court to the Rule 56.1 Statement of Undisputed Facts for a full recitation of the material facts in this case.

I. RULE 56 STANDARDS

Plaintiffs-Intervenors meet the Rule 56 requirements for the grant of summary judgment. The moving party has the initial burden of showing that the requisites of Rule 56(c) are met. *Celotex Corp. v. Catrett*, 477 U.S. 317, 323 (1986). Once a moving party meets its Rule 56(c) burden, the non-movant "may not rely merely on allegations or denials in its own pleading," Rule 56(e), but must "set forth specific facts showing that there is a genuine issue for trial." *Anderson v. Liberty Lobby*, 477 U.S. 242, 256 (1986); *see also SEC v. Research Automation Corp.*, 585 F.2d 31, 33 (2d Cir. 1978). The non-movant "must do more than simply show that there is some metaphysical doubt as to the material facts." *Matsushita Elec. Indus. Co. v. Zenith*

Radio Corp., 475 U.S. 574, 586 (1986)(internal quotations omitted). It cannot resist summary judgment through "mere speculation or conjecture as to the true nature of the facts," *Lipton v. Nature Co.*, 71 F.3d 464, 469 (2d Cir. 1995)(citations omitted), or through affidavits amounting to "self-serving conclusions." *BellSouth Telecomm. v. W.R. Grace*, 77 F.3d 603, 615 (2d Cir. 1996).

When the motion for summary judgment involves issues upon which the non-movant carries the burden of proof at trial, the moving party meets its Rule 56 burden by "pointing out to the district court that there is an absence of evidence to support the nonmoving party's case." *Celotex*, 477 U.S. at 325. Nothing more is required of the moving party in such circumstances:

> [T]he plain language of Rule 56(c) mandates the entry of summary judgment, after adequate time for discovery and upon motion, against a party who fails to make a showing sufficient to establish the existence of an element essential to that party's case, and on which that party will bear the burden of proof at trial. In such a situation, there can be 'no genuine issue as to any material fact,' since a complete failure of proof concerning an essential element of the nonmoving party's case necessarily renders all other facts immaterial.

Celotex, 477 U.S. 317-318.

This principle applies when expert opinion is necessary to establish an element of the non-movant's case. Summary judgment against that party is appropriate when the expert fails to provide either the necessary opinion or the underlying facts upon which that opinion is based. *Miller v. Pfizer, Inc.*, 356 F.3d 1326, 1335 (10th Cir. 2004)(in case involving assertion that Zoloft caused suicide of teenager, summary judgment granted to defendants as plaintiffs had failed to file timely supplemental report, and when report was excluded on *Daubert* grounds, plaintiffs were without scientific evidence of causation—an element on which they bore the burden of proof); *Firefighter's Inst. for Racial Equality ex rel. Anderson v. City of St. Louis*, 220 F.3d 898, 904 (8th Cir. 2000) (affirming grant of summary judgment where non-moving party failed to

comply with filing deadline for expert report and therefore lacked basis to challenge defendant's expert testimony on job-relatedness of fire battalion chief exam); *Duplantis v. Shell Offshore, Inc.*, 948 F.2d 187, 191 (5th Cir. 1991)(summary judgment affirmed where non-moving party's expert failed to offer testimony on whether placement of crane cover was unsafe or substandard, an opinion critical to the non-movant's case); *Monolithic Power Sys., Inc. v. O2 Micro Int'l Ltd.*, 476 F.Supp.2d 1143, 1155 (N.D. Cal. 2007)(summary judgment granted where non-moving party failed to obtain underlying evidence that would permit expert to calculate royalty damages—a necessary element of the non-movant's case); *New Mexico v. General Elec. Co.*, 322 F.Supp.2d 1237, 1256 (D.N.M. 2004)(in public nuisance case arising from hazardous chemical contamination of groundwater, summary judgment entered for defendant in light of plaintiff's expert's failure to testify to the existence of a contaminant plume).

II. THERE IS NO GENUINE ISSUE AS TO ADVERSE IMPACT, AND PLAINTIFFS-INTERVENORS ARE ENTITLED TO JUDGMENT AS A MATTER OF LAW

To make out a *prima facie* case of disparate impact under Title VII, a plaintiff must "identify[] the specific employment practice that is challenged" and then "offer statistical evidence of a kind and degree sufficient to show that the practice in questions has caused the exclusion of applicants for jobs or promotions because of their membership in a protected group." *Watson v. Forth Worth Bank & Trust*, 487 U.S. 977, 994 (1987); *see also Robinson v. Metro-North Commuter R.R. Co.*, 267 F.3d 147, 160 (2d Cir. 2001)(same). The statistical disparities "must be sufficiently substantial to raise an inference of causation." *Smith v. Xerox Co.*, 196 F.3d 358, 365 (2d Cir. 1999) (citing *Watson*, 487 U.S. at 994-995); *see also Robinson,* 267 F.3d at 160 (same).

The Second Circuit recognizes standard deviation analysis as the primary statistical methodology for proving adverse impact under Title VII. *See Guardians Ass'n v. Civil Serv.*

Comm'n of New York, 630 F.2d 79, 86 (2d Cir. 1980) (citing *Castaneda v. Partida*, 430 U.S. 482, at 496 n.17 (1977); *Smith v. Xerox Corp.*, 196 F.3d at 366. Under this approach, in cases involving large sample sizes,[1] a plaintiff can make out a *prima facie* case of disparate impact by showing, through the application of appropriate statistical measures to the test scores of different demographic groups, that "the difference between the expected value (from a random selection) and the observed number is greater than two or three standard deviations." *Guardians*, 630 F.2d at 86, quoting *Castaneda*, 430 U.S. at 496, n.17.

In this case, Plaintiffs-Intervenors have challenged four (4) specific employment practices: Defendants' use of written Exams 7029 and 2043 as pass-fail screening devices and in the rank-order processing of candidates. (Intervenors' Compl., Dkt. 48 at ¶¶ 52-55). In all four cases, the statistical significance of the disparity between white and black firefighter candidates has been established by Plaintiffs-Intervenors' and Plaintiff's experts, and has been admitted by Defendants or their experts. Plaintiff's expert's calculations of "practical significance" – the actual impact of the challenged exams on applicants – have not been disputed.

A. THE DISPARITIES IN THE PASS RATES AND ELIGIBILITY LIST RANKINGS OF BLACK AND WHITE CANDIDATES ON WRITTEN EXAMS 7029 AND 2043 ARE STATISTICALLY SIGNIFICANT

1. Disparity in Pass Rates on Exam 7029

Plaintiff's and Plaintiffs-Intervenors' experts found, and Defendants themselves admit, that the disparity in the pass rates of white and black candidates on Written Exam 7029 exceeded three (3) units of standard deviation.[2] (App. O:115,121-122, 132, Siskin I at 3, 21-22, Table 1;

[1] Exam 7029 had 12,915 white and 1,749 black applicants (totaling 14,664), and Exam 2043 had 13,878 white and 1,393 black applicants (totaling 15,271). (Plaintiffs-Intervenors' Rule 56.1 Statement of Facts, hereinafter referred to as "SOF" at ¶¶ 27, 32).

[2] Plaintiffs-Intervenors' and Plaintiff's experts actually calculated the disparity to be more than thirty-three (33) units of standard deviation. (App O: 121, 132; App R:155). This figure and all

App. R:154-155, Wiesen I at 18-19; App. M:74, Adm. 1).[3] Given the large sample sizes of 12,915 white and 1,749 black test takers (SOF ¶27), this is clearly a statistically significant disparity under the standard set forth in *Guardians*, 630 F.2d at 86 (if the difference "is *greater than two or three* standard deviations, a prima facie case is established")(emphasis added)(citation omitted).

2. Disparity in Pass Rates on Exam 2043

Like Exam 7029, the sample sizes involved in Exam 2043 – 13,878 white candidates and 1,393 black candidates – were quite large. (SOF ¶32). Plaintiffs-Intervenors' and Plaintiff's experts found, and Defendants themselves admit, that the disparity in the pass rates of black and white candidates exceeded three (3) units of standard deviation. (App. M:75, Adm. 3; App. O:117, 126, 135, Siskin I at 5, 26, Table 5; App. R:158-159, Wiesen I at 42-43).[4]

3. Disparity in Eligibility List Rankings of Black and White Candidates on Exam 7029

Defendants admit that, as a group, black candidates who passed Exam 7029 were ranked lower on the Exam 7029 eligibility list than were white candidates. (App. M:76, 78, Adm. 5, 9). Plaintiffs-Intervenors' and Plaintiff's experts calculated, and Defendants themselves admit, that

---

of the standard deviation figures set forth in footnotes 4-7 equate to a less than 1 in 10,000 probability that the disparities between white and black candidates in passing and ranking occurred by chance. (App O:121-122, 124, 126, 128-130).

[3] Citations to the Appendix refer to the lettered exhibit tab where the document may be found as well as the Appendix page number, at the upper right-hand corner of each page. For example, the first page of the Appendix is "App. A:1" and the final page is "App. RR:590". Deposition transcripts are referred to by their location in the Appendix, the name of the witness as well as the transcript page, e.g., "App. T:231, Bobko Tr. 319." Defendants responses and objections to Plaintiff's and Plaintiffs-Intervenors' requests for admission are referred to by location in the Appendix and admission number, e.g., "App. M:74, Adm. 1." Likewise, Defendants' responses and objections to Plaintiff's and Plaintiffs-Intervenors' interrogatories are referred to by location in the Appendix and the interrogatory number, e.g., "App. N:107, Interrog. 30."

[4] Plaintiffs-Intervenors' and Plaintiff's experts actually calculated the disparity to be more than twenty-one (21) units of standard deviation. (App O:126; App R:159).

the disparity between the average rank of white candidates and the average rank of black candidates on the eligibility list for Exam 7029 exceeded three (3) units of standard deviation. (App. M:76-79, Adm. 5, 6, 9, 10; App. O:124, Siskin I at 24; App. R: 156, Wiesen I at 27).[5] Again, because this disparity involves large samples, it is clearly statistically significant. *See Guardians*, 630 F.3d at 86.

4. Disparity in Eligibility List Rankings of Black and White Candidates on Exam 2043

Plaintiff's and Plaintiffs-Intervenors' found, and Defendants' experts admit, that as a group black candidates who passed Exam 2043 were ranked lower on the eligibility list for Exam 2043 than were white candidates, and that the disparity between the average rank of white candidates and the average rank of black candidates on the eligibility list for Exam 2043 was at least 2.99 units of standard deviation. (App. O:128-129, Siskin I at 31-32; App. R:160-161, Wiesen I at 51-52; App. S:170-171, Bobko-Schemmer at 17-18).[6] In addition, Plaintiff's and Defendants' experts both concluded that the disparity between the proportion of black and white candidates on the Exam 2043 eligibility list ranked at or above the last person appointed from that list exceeded three (3) units of standard deviation. (App. O:129-130, Siskin I at 32-33; App. S:170, Bobko-Schemmer at 17).[7] Again, given the large sample size, these standard deviation calculations are statistically significant. *See Guardians*, 630 F.3d at 86.

---

[5] Plaintiffs-Intervenors' and Plaintiff's experts actually calculated the disparity to be more than six (6) units of standard deviation. (App O:124; App R:156).

[6] Plaintiff's and Plaintiffs-Intervenors' experts actually calculated the disparity to be more than nine (9) units of standard deviation. (App O:128-29; App. R:160-61).

[7] Plaintiff's expert actually calculated the disparity to be more than nine (9) units of standard deviation. (App O:129-130).

B. DEFENDANTS' SAMPLE-SIZE ARGUMENT DOES NOT ELIMINATE ADVERSE IMPACT

Defendants' experts make the bizarre argument that the very high levels of disparity found by Plaintiffs-Intervenors' and Plaintiff's experts are artificially inflated because the sample sizes were too big. (App. S:168-169, 177-179, Bobko-Schemmer at 14-15, Appendix A) Defendants' argument is peculiar because it is *small* rather than large sample sizes which the Second Circuit has held to undermine the reliability of statistical significance testing. *See*, *e.g.*, *Waisome*, 948 F.2d at 1379; *Lowe v. Commack Union Free School Dist.*, 886 F.2d 1364, 1372 (2d Cir. 1989). Furthermore, the Court in *Guardians* applied the two-to-three standard deviation analysis to an applicant pool of 36,797 – much larger than the applicant pool involved with Exams 7029 and 2043. 630 F.2d at 85. Nevertheless, Plaintiff's expert calculated, and Defendants' experts concede, that even if the sample sizes were reduced by 90%, the disparities between black and white candidates' pass rates and eligibility list rankings on Exams 7029 and 2043 would *still* exceed two (2) or three (3) units of standard deviation. (App. O:122, 124, 126-127, Siskin I at 22, 24, 26-27, App. S:170-172, Bobko-Schemmer at 17-19).

C. THE DISPARITIES IN THE PASS RATES AND ELIGIBILITY LIST RANKINGS OF BLACK AND WHITE CANDIDATES ON BOTH EXAMS ALSO HAD PRACTICAL SIGNIFICANCE

1. Practical Significance of the Disparities in Pass Rates

Plaintiffs-Intervenors' and Plaintiff's experts both calculated that the passing shortfalls for black candidates on Written Exams 7029 and 2043 – that is, the number of additional black candidates who would have passed had the black pass rate been equal to the white pass rate – were at least 457 for Exam 7029 and 150 for Exam 2043. (App. O:122, 127, Siskin I at 22, 27; App.R:155, 159, Wiesen I at 19, 43). Defendants' experts did not contest Plaintiff's expert's calculations under this methodology (App. T:197-200, Bobko Tr. 139, 142-143, 145), and this

and similar shortfall methodologies have been recognized by federal courts as legitimate measures of practical significance in Title VII disparate impact cases. *See, e.g., Delgado v. Ashcroft*, No. 99-2311, 2003 U.S. Dist. LEXIS 26471, *24 (D.D.C. May 29, 2003); *Waisome*, 948 F.2d at 1376-77. Shortfalls of 457 and 150 passers are substantial numbers, the former being greater than the total number of black firefighters currently employed by the FDNY and the latter roughly equal to half the size of a typical class of new firefighter hires in the FDNY's fire academy. These passing shortfalls are "hardly *de minimis*" and are large enough to establish practical significance. *Delgado*, 2003 U.S. Dist. LEXIS 26471 at *24 (shortfall of 45 Hispanics established practical significance).

2. Practical Significance of the Disparities in Eligibility List Rankings

Plaintiff's expert Dr. Siskin used similar methodologies to calculate the practical significance of the disparities between black and white candidates' rankings on the eligibility lists for Exams 7029 and 2043. First, hypothesizing what the hiring dates for black candidates would have been had their average ranking distribution (and thus appointment date distribution) on the Exam 7029 eligibility list been the same as the average ranking and appointment date distribution for white candidates, Dr. Siskin calculated that 68 black candidates from the Exam 7029 eligibility list had suffered delays in their appointments to the FDNY.[8] (App. O:116, 125, 133-134, Siskin I at 4, 25, Tables 3A, 3B). For Exam 2043, Dr. Siskin calculated that approximately 30 to 40 additional black candidates would have been appointed if the proportion of black test passers ranked above the last person appointed had been equal to the proportion of

[8] Because the FDNY hired enough firefighters to exhaust the entire Exam 7029 eligibility, test passers' low rankings delayed their appointments (often by several years), but did not prevent them. However, this particular analysis does not take into account the additional 457 black test-takers who would have been placed on the Exam 7029 eligibility list, and thus eligible for hire, but for the use of a cutoff score of 84.705 on Written Exam 7029.

white test passers ranked above the last person appointed off of that list. (App. O:129-130, 136, Siskin I at 32-33, Table 11).[9]

Finally, Dr. Siskin calculated that 44 black candidates on the Exam 2043 eligibility list were *delayed* in their appointment to the FDNY because of the disparities in the average eligibility list rankings of black and white candidates. (App. O:130-131, 138, Siskin I at 33-34, Table 12B). Plaintiff-Intervenor Candido Nuñez was one of these delayed hires. Nuñez sat for Written Exam 2043 in December 2002 but was not appointed for more than five (5) years, in the last class from that list in January 2008. (App. I:47; App. TT:599-600, Nuñez Tr. 58-59).

Again, Defendants' experts did not contest any of Dr. Siskin's calculations based on these methodologies, (App. T:197-204, Bobko Tr. 139, 142-143, 153-156), and all three shortfall numbers are clearly more than "*de minimis*." Accordingly, as a matter of law, the disparities in the rankings of black and white candidates on Exams 7029 and 2043 had practical significance.

D. THE "80 % GUIDELINE" DOES NOT DEFEAT PLAINTIFFS-INTERVENORS' PRIMA FACIE CASE

Unable to refute Plaintiffs-Intervenors' and Plaintiff's overwhelming evidence of statistical and practical significance, Defendants' experts argue that some of the disparities in the pass rates and eligibility list rankings of black and white candidates on Exams 7029 and 2043 do not satisfy the so-called "80%" or "Four-fifths rule" for establishing adverse impact under Title VII. (App. S:164-167, Bobko-Schemmer at 8-11). However, application of the rule to the present case is of no help to Defendants.

Under the 80% test, which is set forth in the United States Equal Employment Opportunity Commission's *Uniform Guidelines on Employee Selection Procedures* ("*Uniform Guidelines*"), 29 C.F.R. § 1607.4 (1978), "[a] selection rate for any race, sex, or ethnic groups

[9] Again, this shortfall figure does not include the shortfall caused by the cutoff score.

which is less than four-fifths (4/5) (or eighty percent) of the rate of the group with the highest rate will generally be regarded [] as evidence of adverse impact." 29 C.F.R. §1607.4D.[10] Applying this guideline to the black and white pass rates on Written Exam 7029 supports a finding of adverse impact. Defendants admit that the ratio of the pass rate of black candidates to the pass rate of white candidates was less than 80%. (App. M:80, Adm. 13).

Similarly, application of the 80% guideline to the black-white disparity in eligibility list ranking for Exam 2043 supports a finding of adverse impact because, as Defendants admit, the proportion of black candidates on the Exam 2043 eligibility list who ranked at or above the last person appointed from that list was less than 80% of the proportion of white candidates on the eligibility list who ranked at or above the last person appointed. (App. M:81-82, Adm. 16-17).

As for the disparity between the average eligibility list rankings of black and white firefighter candidates on Exam 7029, neither Plaintiff's, Plaintiffs-Intervenors' nor Defendants' experts determined whether or not the disparity was within the 80% guideline.

---

[10] Notably, the 80% approach itself is nothing more than a non-binding "rule of thumb for the courts." *Watson*, 487 U.S. at 995 n.3; *EEOC v. Joint Apprenticeship Comm.*, 186 F.3d 110, 118 (2d Cir. 1998)(same); *see also Guardians*, 630 F.2d at 91 (discussing non-binding nature of the *Uniform Guidelines*). More importantly, as both Defendants' expert Dr. Bobko and the EEOC have acknowledged, the 80% guideline is a *less* reliable indicator of adverse impact than statistical significance testing in cases, like this one, involving large sample sizes. (App. T:191-194, Bobko Tr. 61-62, 65-66; *Uniform Guidelines Q's & A's*, 44 Fed. Reg. 11996, 11999 (March 2, 1979) at 20, 22, available at www.eeoc.gov/policy/qanda_clarify_procedures.html). As the *Uniform Guidelines* note, between-group selection rate differences that may not satisfy the 80% guideline's definition of adverse impact "may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms," 29 C.F.R. § 1607.4D, a situation which the EEOC acknowledges is most likely to occur in cases involving large sample sizes. *See Uniform Guidelines Q's & A's*, 44 Fed. Reg. 11996, 11999 at 20, 22. The Second Circuit has found a *prima facie* case of adverse impact without even considering evidence related to the 80% test where standard deviation analysis showed clear statistical significance. *See, e.g., Bridgeport Guardians v. Bridgeport*, 933 F.2d 1140, 1146-47 (2d Cir. 1991); *Joint Apprenticeship Comm.*, 186 F.3d at 118-120.

Defendants will likely note that, on Written Exam 2043, the black pass rate was 88% of the white pass rate. However, the disparity between black and white pass rates amounted to an astounding twenty-one (21) units of standard deviation, and Defendants and their experts admit that, even after reducing the sample sizes by 90%, the disparity *still* exceeded three (3) units of standard deviation. (App. M:75, Adm. 3; App. O:117, 126, 135, Siskin I at 5, 26, Table 5; App. R:158-159, Wiesen I at 42-43; App. S:171, Bobko-Schemmer at 18). In other words, there is a statistically significant showing of adverse impact.

The decision in *Delgado v. Ashcroft* is particularly instructive here. As in the case at bar, the *Delgado* plaintiffs established that the use of the challenged employment practice resulted in a disparity between Hispanics and whites that was both statistically significant, as measured by the number of standard deviations, and practically significant, as measured by a hiring shortfall of 45 Hispanics over a five-year period when the challenged practice was used. 2003 U.S. Dist. LEXIS 26471 at *8-11. The district court found that plaintiffs had established a *prima facie* case of adverse impact even though the selection rate for Hispanic applicants was 88% of the selection rate for whites. Id. at *24-26. Here, given the strong evidence of statistical significance, an even more substantial passing shortfall of 150 black candidates, and an identical 88% black-to-white pass rate ratio, Written Exam 2043 clearly had an adverse impact as well. *See also Graves v. Alabama State Bd. of Educ.*, 776 F.Supp. 1518, 1528 (M.D. Ala. 1991) (finding adverse impact where disparity between black and white test takers had "overwhelming statistical significance" although "the result of analysis under the four-fifths rule [wa]s at most a borderline figure").

E. DEFENDANTS' EXPERT'S ADMISSIONS AND FAILURES TO OPINE ARE DISPOSITIVE

In addition to the admitted and undisputed facts described above, Defendant's expert Dr. Philip Bobko explicitly admitted at deposition that both the use of Written Exam 7029 as a pass/fail screening device with a cut-off score of 84.705 and the rank-order processing of firefighter candidates from the Exam 7029 eligibility list had an adverse impact on black firefighter candidates. (App. T:185, 189, Bobko Tr. 52, 56). Dr. Bobko's admission eliminates any disputed issue of fact as to the adverse impact of Exam 7029.

When asked for his opinion as to the adverse impact on black candidates of the use of Written Exam 2043 as a pass/fail screening device and for rank order processing of candidates, Dr. Bobko *refused* to offer one. (App. T:187-188, 190, Bobko Tr. 54-55, 57). Given Plaintiffs-Intervenors' and Plaintiff's experts' unequivocal and factually supported opinions that these employment practices had an adverse impact on black candidates (App. O:118, Siskin I at 10; App. R:159, 161, Wiesen I at 43, 52), Dr. Bobko's failure to offer an opposing opinion eliminates any dispute of fact as to the first prong of Plaintiffs-Intervenors' disparate impact claims.[11]

III. THERE IS NO GENUINE ISSUE OF FACT AS TO JOB-RELATEDNESS AND BUSINESS NECESSITY, AND PLAINTIFFS-INTERVENORS ARE ENTITLED TO JUDGMENT AS A MATTER OF LAW

In light of this showing of adverse impact, the burden shifts to Defendants to prove that the two challenged examinations are job related and consistent with business necessity. *Gulino*, 460 F.3d 361, 382 (2d Cir. 2006); *Hack v. President & Fellows of Yale Coll.*, 237 F.3d 81, 100

[11] The reason Dr. Bobko gave for not opining on these issues, that he needed "additional clarification" on the meaning of the term "adverse impact" (App. T:187-188, 190, Bobko Tr. 54-55, 57) was strange considering that, only a few minutes earlier, he was able to offer an opinion as to the adverse impact of using Exam 7029 as a pass/fail screening device without seeking any additional clarification. (App. T:185, Bobko Tr. 52).

(2d Cir. 2000)("Congress has required a Title VII defendant to justify not only the legitimacy of the ends, but also the necessity of the means"), citing *Lanning v. SEPTA*, 181 F.3d 478, 488 (3d Cir. 1999); *Nash v. Consol. City of Jacksonville*, 895 F. Supp. 1536, 1545 (M.D.Fla. 1995) ("The Civil Rights Act of 1991... restored pre-Wards Cove law in which, *inter alia*, ... 'business necessity' really means 'necessity.'"), aff'd, 85 F.3d 643 (11th Cir. 1996).

Under *Guardians*, an employment test with adverse impact violates the second prong of Title VII if *either* the content of the exam is not validated *or* the uses to which it is put have not been validated. 630 F.2d at 87. Regarding the police officer exam at issue in *Guardians,* the court held that "even if the construction of the exam passes muster, *the way in which it was used* to distinguish among candidates . . . defeats any claim of validity for a testing process that produces disparate racial results." 630 F.2d at 99-100 (emphasis added). There, "the City *used* the results of the exam to compile a rank-ordering of all the applicants, and then selected a passing score sufficient to generate the required number of potential trainees." 630 F.2d at 100 (emphasis added). The Court flatly rejected those uses of the exam because the City had failed to show that the uses were related to job performance. 630 F.2d at 105-106 ("Primarily on the basis of [the challenged exam's] improper use of rank-ordering, and of the cutoff score, we affirm the conclusion of the District Court that the exam as used was invalid") .

Defendants here *used* the exams in the same ways struck down by *Guardians*: (1) to set pass/fail cutoff scores (84.705 on Exam 7029 and 70.000 on Exam 2043) that eliminated all candidates who scored below those marks, and (2) to rank-order the remaining candidates based on their exam scores for the purpose of hiring. Defendants have failed to present any evidence to support the validity of either of these uses. Therefore, Defendants cannot meet their burden to show that the exams are lawful under Title VII.

A. EXAMS 7029 AND 2043 ARE NOT VALID BECAUSE THE CUTOFF SCORES FOR PASSING WERE NOT VALIDATED

The Second Circuit in *Guardians* held that "when an exam produces disparate racial results, a cutoff score requires adequate justification," 630 F.2d at 106, "so that, for any given cutoff . . . those who passed would likely perform the job better than those who failed." 630 F.2d at 105. Here, Defendants made no attempt to present evidence justifying the cutoff scores used on Exams 7029 and 2043. Defendants' expert Dr. Bobko conceded this:

> Q. In the parts of your report where you talk about job relatedness and business necessity, you didn't discuss the pass/fail cutoff scores the City used on written exam 7029 and 2043, correct?
>
> A. [Bobko:] Correct.

(App. T:231, Bobko Tr. 319).

When asked directly whether his and Dr. Schemmer's report established the validity of the cutoff scores used, Dr. Bobko admitted that it did not:

> Q. Is what's in this report, the Bobko, Schemmer report, sufficient to establish that the City's use of exam 7029 as a pass/fail screening device with a cutoff point of 84.705 is consistent with job relatedness and business necessity?
>
> A. [Dr. Bobko:] No.
>
> Q. Is it your opinion it is sufficient to establish that the City's use of written exam 2043 as a pass/fail screening device with a cutoff score of 70 is job related and consistent with business necessity?
>
> A. [Dr. Bobko:] No.

(App. T:205-206, Bobko Tr. 179–180).

The law is clear that where a defendant fails to demonstrate that a one- or two-point difference in scores reflects a meaningful difference in performance, and where the exam has an adverse impact, the defendant cannot set a cutoff score solely on the basis of the expected

number of applicants needed to fill vacancies. *Guardians*, 630 F.2d at 105. Yet that is precisely what Defendants did, as explicitly and consistently admitted by their witnesses.

Carol Wachter, the Assistant Commissioner for Examinations at the Department of Citywide Administrative Services ("DCAS") when Exam 7029 was developed, testified that no validity study was conducted with respect to the pass mark for Written Exam 7029. (App. W:299, Wachter Tr. 85). Rather, the pass mark was set based on the number of entry-level firefighter job openings anticipated by the FDNY. (App. W:297-298, 303-304, Wachter Tr. 74-75, 181-82). Wachter also testified that it would have been possible to use a lower pass mark for Written Exam 7029 and still select only qualified candidates. (App. W:301-302, Wachter Tr. 179-180). In fact, she agreed that a pass mark of 84.705 was in excess of the minimum competency level necessary for the job. (App. W:300, Wachter Tr. 86).

Thomas Patitucci, who took over Ms. Wachter's position, also testified that the Written Exam 7029 cutoff score was not based on data from the job analysis but rather was based on the number of candidates that the FDNY wished to have pass the written test and be called to the physical. (App. P:144-146, Patitucci (designee) Tr. 91-93). Patitucci characterized the pass mark for Written Exam 7029 as "unrealistically high." (App. P:141-143, Patitucci Tr. 42-44).

The cut-off score for Exam 2043 was set pursuant to New York City's Personnel Rules and Regulations 4.4.9(a), which provides that unless otherwise specified by the Notice of Exam or the DCAS Commissioner, candidates must score at least 70% to be placed on an eligibility list. (App. N:105-106, Interrog. 16). The cutoff was not based on meaningful differences in levels of job performance, as required by *Guardians*, 630 F.2d at 105.

Dr. Siskin highlights one of the problematic effects of Defendants' inconsistent use of cutoff scores. If a hypothetical Applicant "A" scored 84.705 on Written Exam 7029 and 75 on

the physical performance test ("PPT"), s/he would have had a transformed score of 70.000, under the methodology for calculating transformed scores set forth in the "Firefighter Exam No. 7029 Explanation of Test Scores." (App. Y:328-329; App. V:283, Siskin II at 15). If a hypothetical Applicant "B" scored 83.529 on Written Exam 7029 and 100 on the PPT (and a majority of PPT scores were 100, App. Z:341, Wiesen II at 82), s/he would have had a transformed score of 83.652. Id. Applicant B would have a considerably higher transformed score than Applicant A and should be higher on the eligibility list. The problem is that Applicant B's score on the written test would have disqualified him or her from even taking the PPT. (App. M:95, Adm. 68). Thus, under the City's use of Exam 7029, Applicant A (with a transformed score of 70.000) would be placed on the Exam 7029 eligibility list, but Applicant B (with a transformed score of 83.652) would not. This potentially prevented Defendants from selecting the most qualified candidates for the job. Moreover, it gave unwarranted weight to the discriminatory written portion of the test, in spite of Defendants' intention to weigh the written and physical components of the exam equally.

Guardians puts the burden squarely on Defendants to establish that the cutoff scores used on the exams were valid. 630 F.2d at 105-106. Despite this, they have put forward no expert opinion or other evidence to satisfy that burden. The do not – and cannot – show that there is a meaningful difference in the abilities of those who passed the exams and those who failed. The well-supported opinions of Plaintiff's and Plaintiffs-Intervenors' experts that the exams were not valid as used is therefore uncontroverted. (App. Z:341-342, Wiesen II at 82-83; App. V:283-290, Siskin II at 15-22). Defendants have not and cannot meet their burden on this critical issue of "cutoff score"; for this reason alone judgment for Plaintiffs-Intervenors is required in this case.

B. EXAMS 7029 AND 2043 ARE NOT VALID BECAUSE THEIR RANK-ORDER PROCESSING OF CANDIDATES WAS NOT VALIDATED

For rank-ordering of candidates to be a permissible use of an exam, there must be "a demonstration of such substantial test validity that it is reasonable to expect one- or two-point differences in scores to reflect differences in job performance." *Guardians*, 630 F.2d at 100-101. The City could not show this to be the case in *Guardians*, and it cannot do so here.

Defendants' expert report on job relatedness and business necessity, which was three (3) pages long, never stated *any* opinion as to the whether the rank-order processing of candidates based on their exam scores was valid. (App. S:173-175, Bobko-Schemmer at 27-29). When questioned about the report, one of its co-authors Dr. Bobko testified as follows:

> Q. Is it your opinion that what is in this report is sufficient to establish that the City's rank/order/processing and selection of candidates from the exam 7029 eligibility list is job related and consistent with business necessity?
>
> A. [Dr. Bobko:] No.
>
> Q. Is it your opinion that what is in this report is sufficient to establish that the City's rank/order/processing and selection of candidates from the exam 2043 eligibility list is job related and consistent with business necessity?
>
> A. [Dr. Bobko:] No.

(App. T:205-206, Bobko Tr. 179–180).

In spite of their clear burden of proof, Defendants' experts did not even attempt to show a relationship (let alone a close relationship) between a candidate's ranking on the eligibility list and his or her ability to perform the job. The weight of this failure cannot be overcome. As the Second Circuit stated in *Guardians*, 630 F.2d at 89, and re-iterated in *Gulino*, 460 F.3d at 383, employment testing is not primarily a legal subject, but rather is an area where the courts "must take into account the expertise of test validation professionals." In most cases, the question of

job-relatedness is a question that is dependent upon expert testimony respecting validity studies. As noted in *Gulino*, 460 F.3d at 383: "the 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" (internal cites omitted). Defendants' failure to assert *any* correlation between ranking and job performance on either challenged exam precludes them from meeting their burden of proving the job relatedness of the tests as used.

In addition, there is affirmative evidence that one- or two-point differences in scores *cannot* predict job performance on these exams. As Dr. Siskin shows, the standard error of measurement for Written Exam 7029 is 2.64 and for Written Exam 2043 is 2.63. (App. V:291, Siskin II at 23). Based on a statistical calculation, this translates into a margin of error of four (4) questions. Id. Thus, on both Written Exams 7029 and 2043, four (4) wrong answers are within the range of normal variation in scores due to *chance* and may not represent any difference in ability between individuals. Id. Yes, four (4) wrong answers can have a great impact on a candidate's rank. On Written Exam 7029, four (4) wrong answers would drop a candidate's maximum possible rank on the eligibility list from 1 to 2,454. Id. Four (4) wrong answers on Written Exam 2043 would drop a candidate's maximum possible rank from 1 to 1,713. Id. Thus, a candidate's rank could easily vary by thousands of places on the hiring list due to *chance alone*, regardless of his or her abilities. As the Court in *Guardians* wisely stated:

> Rank-ordering satisfies a felt need for objectivity, but it does not necessarily select better job performers. In some circumstances the virtues of objectivity may justify the inherent artificiality of the substantively deficient distinctions being made. But when test scores have a disparate racial impact, an employer violates Title VII if he *uses* them in ways that lack significant relationship to job performance.

Guardians, 630 F.2d at 100 (emphasis added).

The Second Circuit has also held that a test may not be used for rank-ordered selections where there is no showing of the test's reliability, i.e. "the extent to which the exam would produce consistent results if applicants repeatedly took it or similar tests." *Guardians*, 630 F.2d at 101. Here, Defendants made no effort to demonstrate that the challenged exams were reliable, and the evidence is clear to the contrary. Many applicants sat for more than one entry-level firefighter exam, providing a good empirical test of the reliability of the examinations. Of the 17,145 applicants who took Written Exam 7029, 2,667 also took Written Exam 2043. (App. U:259, Jones-Hough at 39). Among those who took both written exams, 54.8% of those who scored below 70 on Exam 2043 scored 70 or above on Written Exam 7029. (App. V:287, Siskin II at 19). Among those who scored below 70 on Written Exam 7029, 75.9% scored 70 or above on Written Exam 2043. Id. Scores were so divergent that out of all those who took both written exams and scored below 70 on one of them, 81.3% scored above 70 on the other. Id.

The unreliability of Exams 7029 and 2043 is apparent when comparing them to Exam 6019 as well. Exam 6019 (given in January 2007) evaluated candidates for the same firefighter job as Exams 7029 and 2043. Of those candidates who took Written Exam 6019, 479 had previously taken Written Exam 7029 and 2,240 had previously taken Written Exam 2043. (App. U:259, Jones-Hough at 39). One-hundred fifty-six (156) of the candidates who took Written Exam 6019 had scored below 70 on either Written Exam 7029 or Written Exam 2043, but 84.6% of them passed (scored over 70 on) Written Exam 6019. (App. V:287-288, Siskin II at 19-20).

These numbers demonstrate a total lack of reliability between examinations that were used to select candidates for the very same position. Defendants' experts conducted no reliability analysis, and thus no evidence exists to contradict Plaintiff's expert's findings. As the *Guardians* court held "[w]ithout some substantial demonstration of reliability it is wholly

unwarranted to make hiring decisions, with a disparate racial impact, for thousands of applicants that turn on one-point distinctions among their passing grades." 630 F.2d at 101. Clearly, there has been no "substantial demonstration of reliability" here.

In short, rank-ordering has not been shown – and cannot be shown – to be valid. On this ground alone the challenged exams are plainly unlawful under Title VII and summary judgment in favor of Plaintiffs-Intervenors is warranted.

C. NO NEW EXPERT OPINIONS CAN BE OFFERED NOW

The City may not, at this eleventh hour, introduce new expert opinion regarding validity. Any new report would violate not only the scheduling order in this case, Dkt. 181, (under which both fact and expert discovery concluded nearly three (3) months ago), but also Rule 26. Rule 26(a)(2) requires a party to disclose, as directed by the court, its expert witnesses and a report that "contain[s] a complete statement of all opinions to be expressed and the basis and reasons therefore." Where a party has the burden of proof on an issue – as the Defendants here have the burden of proving the validity of the challenged exams *and their usage* – it may not submit a supplemental expert report on that issue, unless some sort of new information has emerged. *See Allgood v. General Motors Corp.*, 2007 WL 647496, at *6-7 (S.D.Ind. Feb 2, 2007)(prohibiting the submission of a supplemental expert report on an issue that was "in the case from the beginning" and on which that party clearly had the burden of proof, where the delay in submitting the expert's opinion was clearly strategic); *Wechsler v. Hunt Health Sys., Ltd.*, 381 F.Supp.2d 135 (S.D.N.Y. 2003) (granting motion to strike a supplemental report that substantially changed an earlier analysis, save for some sections that legitimately corrected errors in computation); Fed. R. Civ. P. 26, Adv. Comm. Notes ("the party with the burden of proof on an issue should disclose its expert testimony on that issue before other parties are required to make their disclosures with respect to that issue"); *Mabrey v. U.S.*, 2006 WL 1891127, at *4-5

(D. Nev. July 7, 2006); *Baldwin Graphic Sys., Inc. v. Siebert, Inc.*, 2005 WL 1300763, at *2 (N.D.Ill. Feb. 22, 2005)("A party presents its arguments as to the issues for which it has the burden of proof in its initial expert report"). Here, Defendants have long known that the validity (and valid use) of the challenged exams was their central burden, and they had ample time to develop an expert defense. No new information has emerged that would justify supplementing the original expert report.

The scheduling order here did not provide for rebuttal reports. Had the City desired permission to rebut Plaintiff's and Plaintiffs-Intervenors' experts' opinions, it had three (3) months from July 31, 2008 (when the Plaintiffs' expert reports were submitted) until October 31, 2008 (the close of discovery) to seek such permission. Now, long after the close of discovery, with trial imminent, the City cannot introduce new expert opinions. *See O2 Micro Int'l Ltd. v. Monolithic Power Sys., Inc.*, 467 F.3d 1355, 1368 (Fed. Cir. 2006) (upholding the denial of two attempts to supplement an original expert report, citing the lack of an adequate explanation for the failure to include a crucial theory in the original expert report or for waiting months to attempt a supplemental report). Thus, there is nothing Defendants can offer at this point to establish that the cutoff scores and rank-ordering used on Exams 7029 and 2043 were job related or consistent with business necessity.

D. THE CHALLENGED EXAMS DO NOT MEET THE SECOND CIRCUIT'S OTHER REQUIREMENTS FOR CONTENT VALIDITY

The Court can readily decide the second prong of the Title VII disparate impact analysis based entirely upon Defendants' un-validated *use* of rank-ordering and cutoff scores, as discussed above. While there should be no need for the Court to pass upon the considerable

uncontroverted evidence that flaws in test construction destroyed any underlying validity of the challenged exams, several of these fatal flaws are discussed below.[12]

The Second Circuit in *Guardians*, 630 F.2d 91, reaffirmed that employers may establish the validity of a challenged selection device by showing evidence of either: (1) criterion-related validity, (2) construct validity, or (3) content validity. *See also Uniform Guidelines* §1607.5A-B. Under the *Uniform Guidelines*, criterion-related validity is "[d]emonstrated by empirical data showing that the selection procedure is predictive of or significantly correlated with important elements of work behavior." §1607.16F. Construct validity is "[d]emonstrated by data showing that the selection procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important for successful job performance." Id. Content validity is "[d]emonstrated by data showing that the selection procedure is representative of important aspects of performance on the job." Id.

Defendants do not attempt to show either criterion-related or construct validity. (App. CC:394-395, 397, Adm. 20, 23, 62). Instead, they purport to rely on a content validity strategy.[13] Yet Defendants have failed to present evidence that the challenged examinations have the "five attributes of an exam with sufficient content validity to be used notwithstanding its disparate racial impact." *Guardians*, 630 F.2d at 95. In fact, nowhere in Defendants' experts' report, or in their depositions, do they state an affirmative opinion that the challenged exams are valid under a "content validity" theory or any other. As this is the affirmative burden placed squarely on Defendants, their failure to present expert support for content validity is fatal to their case.

---

[12] In reviewing the many defects of the challenged exams, the Court should note that where an examination is developed by an employer "in-house," as was done here, the Second Circuit advises "giv[ing] the resulting test careful scrutiny." *Guardians*, 630 F.2d at 96.

[13] To the extent that Defendants rely on a theory of "validity generalization" in support of Exams 7029 or 2043, no such methodology has been accepted by this Circuit. Further, Defendants arguments regarding validity generalization are deeply unsound. (SOF ¶¶137-146).

In order for a challenged exam to have the five attributes of content validity, a defendant must show that: (1) the test-makers conducted a suitable job analysis; (2) the test-makers used reasonable competence in constructing the test itself; (3) the content of the test is related to the content of the job; (4) the content of the test is representative of the content of the job; and (5) the test was used with a scoring system that usefully selects, from among the applicants, those who can better perform the job.[14] *Guardians*, 630 F.2d at 95.

1. The Tests Were Based on a Flawed Job Analysis

Development of an employment exam begins with a suitable "job analysis." (SOF ¶¶84). Here, a group of incumbent firefighters was asked to fill out a Job Analysis Questionnaire, on which they rated the importance of various tasks and cognitive abilities to the performance of their jobs, on a scale of 1 ("not relevant") to 4 ("critical"). (SOF ¶¶87-90).

Respondents were not asked to restrict their importance ratings to those tasks/abilities required on the *first day* of the job. (SOF ¶91). This is an important oversight because under the *Uniform Guidelines*, a test that is supported by content validity may not measure knowledges, skills or abilities that an employee will be expected to learn on the job. *Guardians*, 630 F.2d at 94; *Uniform Guidelines* §1607.5F, 14C(1), *Uniform Guidelines Q's & A's* at 74. A content valid test may *only* measure abilities that candidates are expected to possess prior to the start of their on-the-job training (in this case the FDNY's lengthy firefighter training academy). (App. DD:407, Goldstein at 12). The developer of Exam 6019, Dr. Catherine Cline, was well-aware of this and her Job Analysis Questionnaire for Exam 6019 *did* ask respondents to rate tasks and abilities according to whether they were needed on "Day 1." (App. DD:407-409, Goldstein at 12-14). The failure to use a Day 1 standard in developing the challenged exams was a fatal flaw.

[14] The *Guardians* court found that the City had not properly validated the cutoff score and rank-ordered use of the police officer test at issue in that case. The exams at issue here fail for the same reasons, as discussed above, and based on all the other "attributes" of content validity.

Further, Defendants based their test questions on those tasks and abilities that were rated above an average of 2.5 (between "somewhat important" and "important"). (SOF ¶¶92). Dr. Goldstein, an expert in content validity, identified this as an additional problem with the Defendants' methodology. (App. DD:409, Goldstein at 14) ("To support a content validity strategy, tasks and abilities must meet more stringent criteria for importance as well as meeting a Day 1 standard"). Dr. Cline agreed, testifying that she would never use tasks or abilities that had been rated as "somewhat important" because she would want anything she tested for to be important. (App. DD:409, Goldstein at 14, citing Cline Tr. 118-119).

After Defendants had compiled lists of tasks and cognitive abilities that attained ratings of 2.5 or above, a second group of incumbents, referred to as the Linking Panel, was asked to rate (on a scale of 0 "not relevant" to 3 "critical") the importance of each cognitive ability to each job task. (SOF ¶¶93-95). The Linking Panel raters were deposed and candidly admitted that they did not understand the meanings of the abilities they were being asked to evaluate, and in some cases did not understand that they were rating whether an ability was required to *perform* the job, as opposed to being useful in the training academy. (SOF ¶¶96).

As a result of these flaws, among others, the developers of Exams 7029 and 2043 were not able to determine which abilities were truly important to the job of entry-level firefighter, and therefore could not design a valid examination for the position. In the opinions of the United States' and Plaintiffs-Intervenors' experts, the flaws in the Defendants job analysis were fatal to the validity of the exams. (App. DD:408-410, Goldstein at 13-15 (Defendants' job analysis is "an unreliable foundation for a claim of content validity"); App. U:258, Jones-Hough at 36 (Defendants' job analysis "provides no basis for reaching a conclusion of content validity"); App. Z:330, Wiesen II at 31 (the "job analysis was seriously flawed in design, construction, and

administration," and it is "not suitable to be used to support the content validity" of the test)). The Defendants did not attempt, at all, to rebut these findings that the job analysis was invalid, and, therefore, they cannot establish test validity.

2. The Formulation Of Questions Was So Incompetent That The Exams Could Not Have Been Validated

In *Guardians*, the Court noted with concern that "[t]he questions were initially framed by police officers, who may have had expertise in identifying tasks involved in their job but were amateurs in the art of test construction." 630 F.2d at 96. The same was true here. Firefighters themselves drafted the questions on Exams 7029 and 2043. (SOF ¶¶123-125). Moreover, in *Guardians*, "the questions, although reviewed, were not tested on a sample population. . . the City did not even perform the minimal sample testing to ensure that the questions were comprehensible and unambiguous." Id. Once more, this holds true for the challenged firefighter exams, which were not tested on a sample population. (SOF ¶¶123-125).

A statistical analysis of these exams revealed that the questions did not even measure the abilities they purported to measure. (SOF ¶¶130-135). As Dr. Siskin explains, "one would expect items [questions] intended to measure the same ability to be more highly correlated with each other than with items intended to measure other abilities." (App. V:277, Siskin II at 5). On Written Exam 7029, for eight (8) of the nine (9) abilities purportedly measured, the questions used to measure those abilities correlated more highly with questions intended to measure *other* abilities than they did with those intended to measure the same ability. (App. V:278, 294, Siskin II at 6, Table 1). For example, the questions intended to measure deductive reasoning correlated more highly with questions intended to measure written expression, written comprehension and spatial orientation than they did with other questions intended to measure deductive reasoning. Id. The same is true for four (4) of the nine (9) abilities purportedly measured by Written Exam

2043. (App. V:278, 295, Siskin II at 6, Table 2). A "factor analysis" – a statistical test that can be used to explain such observed correlations – confirmed that these test questions were *not* in fact measuring distinct ability areas. (App. V:281, Siskin II at 9). This eliminates any claim that the exams were competently constructed or their content represented the content of the job, as required by *Guardians*, 630 F.2d at 96-98. As Defendants' experts performed no factor analysis, they cannot rebut Dr. Siskin's finding, and summary judgment is warranted on this ground.

The Court in *Guardians* also noted that "the reading level of the test should not be pointlessly high." 630 F.2d at 99. Dr. Wiesen conducted a readability study of Written Exams 7029 and 2043 and found that the average reading grade level of the items on both exams was above the 12th grade and that both tests contained many questions that required a college or graduate school level of reading ability. (SOF ¶¶126-128). This is "pointlessly high" for a position that currently requires only a high school degree and six (6) months of work experience. (App. L:72, NOE for Exam 6019). Moreover, as explained by Dr. Wiesen, on "a civil service test, a lower reading grade level is important so that low reading skill does not obscure the ability being tested." (App. Z:337, Wiesen II at 58). For example, a candidate with a high level of ability in the areas of visualization, spatial orientation, and memorization may answer incorrectly the questions measuring those abilities if they are written at a grade level that is inappropriately high. As Defendants did no analysis of reading level, they are unable to rebut Dr. Wiesen's finding that the high reading level of these exams destroyed their validity. (SOF ¶ 128).

3. <u>The Cognitive Abilities Tested Were Not Representative Of The Job</u>

To be content valid, the content of a challenged examination must be a "representative sample of the content of the job." *Guardians*, 630 F.2d at 98, citing *Uniform Guidelines* §1607.14C(4). Exams 7029 and 2043 had an extraordinarily limited test content, including only nine (9) out of the eighteen (18) *cognitive* abilities that had been identified as important to the

job and none of the *non-cognitive* abilities that have been found to predict successful performance in the position of firefighter. (SOF ¶105, 112-118).

The two cognitive abilities that received the highest importance ratings from incumbent firefighters in the job analysis for Exam 7092 were oral comprehension and oral expression. (SOF ¶106). Firefighters are in constant communication with each other and the public, often in moments of extreme stress. But, Defendants chose not to test for either of these oral communication abilities on Exam 7029 or Exam 2043. Id. Hence, the content of the test did not include the two abilities that firefighters reported as most important to their jobs.

Moreover, mechanical ability, a cognitive ability that was found by the U.S. Office of Personnel Management in 1980 to be the cognitive ability most highly correlated with successful performance of the entry-level firefighter job, was not tested by the challenged exams, even though it is easily measurable in a paper-and-pencil, multiple-choice format. (SOF ¶111).

Beyond this inadequate representation of cognitive abilities, Defendants entirely ignored non-cognitive abilities, which have often been found to be more important to the firefighter job than cognitive abilities. The U.S. Civil Service Commission published a firefighter job analysis in 1977 which listed the following abilities and characteristics as *most* important to the firefighter job: responsibility, desire to learn, teamwork, activity (energy), and getting along with people. (SOF ¶159). Each of these most important abilities is non-cognitive, and none were tested on Exams 7029 or 2043. (App. M:84, Adm. 29). Out of the entire list of 20 abilities that the Civil Service Commission included in its recommended test weighting plan, only four (4) are cognitive: Problem-solving Ability, Mechanical Ability, Verbal Skills (both written and verbal) and Math Skills. (SOF ¶160). Of these, DCAS measured only Problem Solving and Written Comprehension on Exams 7029 and 2043. (App. M:84, Adm. 28, 30).

O*Net, a U.S. Department of Labor database of job descriptions, identifies the following abilities as the ten (10) most important for firefighters: dependability, cooperation, attention to detail, self control, stress tolerance, initiative, concern for others, analytical thinking, persistence, and social orientation. (SOF ¶163). The only cognitive ability in the "top 10" (analytical thinking) is also the only one measured by Exams 7029 and 2043. (App. M:84, Adm. 28, 30).

The City's own job analysis for firefighter Exam 6019 – administered in 2007 after this lawsuit was filed – also revealed that cognitive abilities had a relatively low level of importance to the performance of the job. In fact, the importance of cognitive abilities was so moderate that a test-taker could pass Exam 6019 without answering a *single* cognitive ability question correctly, as long as he or she answered each of the non-cognitive questions correctly. (SOF ¶119). Despite this, *no* non-cognitive abilities were tested on the challenged exams, even though they were widely available and had demonstrated validity. (SOF ¶¶147-167).

This non-representative test content was maintained in spite of the fact (well known to the City's exam developers) that black test-takers tend to perform more poorly on tests of cognitive ability than on tests of non-cognitive ability. (App. T:195-196, Bobko Tr. 121-122). In fact, cognitive tests have been found to produce three (3) to five (5) times greater racial differences than other valid predictors of job performance. (App. OO:566-567, Alexander Tr. 349-350). The adverse impact of Exams 7092 and 2043 should have come as no surprise to Defendants. The content of those exams not only decreased their validity, it also critically increased their discriminatory impact on black test takers.

The evidence is clear that at each step, from the seriously flawed job analysis, to the construction of an unreliable test, to the use of unjustified scoring systems, Defendants failed to follow the requirements of the law. They have also failed during the course of this litigation to

present any evidence that would preclude summary judgment in favor of Plaintiffs-Intervenors on any of these dispositive issues.

IV. THERE IS NO GENUINE ISSUE AS TO THE AVAILABILITY OF VALID ALTERNATIVES WITH LESS ADVERSE IMPACT, AND PLAINTIFFS-INTERVENORS ARE ENTITLED TO JUDGMENT AS A MATTER OF LAW

If the Court were to reach the third prong in the Title VII disparate impact analysis, it would be to assess the availability of alternative selection procedures that would have reduced adverse impact while maintaining validity. *Gulino*, 460 F.3d at 382. The evidence shows – and Defendants' witnesses consistently admitted – that Defendants had numerous alternatives available to them in 1999 and 2002, in the form of measures of non-cognitive abilities that would be expected to incrementally *increase* test validity and reduce adverse impact on blacks.

As noted above, the DCAS Examiner for Exam 6019, Robert Alexander, testified that cognitive tests have been found to produce three (3) to five (5) times greater racial differences than other valid predictors of job performance such as biodata, personality inventories and structured interviews. (App. OO:566-567, Alexander Tr. 349-350). Moreover, non-cognitive ability tests with demonstrated validity have been available since the 1970s and 80s. (App. OO:568-572, Alexander Tr. 353-357). This testimony is supported by compelling documentary evidence: The Civil Service Commission's job analysis of the firefighter job (which, as discussed above, found non-cognitive abilities to be so important to job performance) included an appendix listing various non-cognitive ability tests that existed as of 1977, and for which there was already evidence of empirical validity. (App. PP:578-582).

According to Dr. Cline, who developed Exam 6019, the field of industrial/organizational psychology began taking personality measures seriously as a predictor of job performance in or around 1993. (SOF ¶152). This means that a test measuring non-cognitive abilities could

certainly have been used by Defendants in 1999 or 2002 to reduce adverse impact while maintaining or adding to the validity of the testing process.

Defendants' own expert, Dr. Schemmer, testified that tests for resistance to stress, teamwork, responsibility, desire to learn, getting along with others and honesty, among others, were available when Exams 7092 and 2043 were used, and that these abilities are important to the job of firefighter. (App. AA:351, Schemmer Tr. 291). Dr. Schemmer also testified that each of the important non-cognitive abilities listed on the Department of Labor's O*Net site were testable using a paper-and-pencil exam when Exams 7029 and 2043 were given. He testified that they "would certainly have the potential" to add to the validity of the exam (App. AA:356-358, Schemmer Tr. 299-301) and that, "[b]roadly or generally speaking, I would anticipate they would have lesser impact than cognitive assessments." (App. AA:359, Schemmer Tr. at 302).

In addition, Defendants' expert Dr. Bobko has an advisory relationship with a testing company called Previsor that develops and markets written tests for, *inter alia*, the following non-cognitive abilities: emotional control, stress tolerance, self confidence, adaptability, independence, responsibility, initiative, social awareness and empathy. (App. RR; App. T:236-237, Bobko Tr. 459-460). Dr. Schemmer agreed that each of the abilities that Previsor's exams test for could have been tested using paper-and-pencil tests at the time Exams 7029 and 2043 were given, and he had no reason to believe Previsor's tests for these abilities are not valid predictors. (App. AA:353-354, Schemmer Tr. 293-94). Dr. Schemmer agreed that well-developed tests of these abilities would have the potential to add to the validity of an exam for firefighter and, according to the literature, would lead to a smaller disparity in black-white results than purely cognitive tests. (App. AA:355-356, Schemmer Tr. 298-99).

The improvements achieved by Exam 6019 were largely due to its inclusion of items called "situational judgment exercises," which were used to measure the following non-cognitive abilities that are important to the job: tolerance for stress, adaptability, tenacity, integrity, work standards, resilience, coordination, and establishing and maintaining interpersonal relationships. (App. L:72). As Dr. Cline admits, the "situational judgment exercises" used on Exam 6019 had been shown to be valid ten (10) years ago. (SOF ¶ 155, App. BB:388-390, Cline Tr. 507-509).

Given that so many valid alternative measures were being used as early as the 1970s and 1980s, Defendants cannot claim that they were unavailable in 1999 or 2002. In fact, Exam 6019 itself – which Defendants admit is at least as valid as Exams 7029 and 2043 and led to higher pass rates among black applicants (App. CC:400-402, Adm. 70-77) – could have been used.

CONCLUSION

Based upon uncontradicted statistical evidence, and as a matter of law, Plaintiffs-Intervenors have established a *prima facie* case of disparate impact under Title VII. Defendants, in response, have utterly failed to meet their burden of establishing that the challenged selection procedures, as used, were job-related or consistent with business necessity. Moreover, Plaintiffs-Intervenors have presented uncontradicted evidence that alternative methods with equal validity and less adverse impact were available at the time that the challenged examinations were used. Therefore, as a matter of law, Plaintiffs-Intervenors have established a violation of Title VII and summary judgment in their favor should be granted.

Dated: February 2, 2009
New York, New York

LEVY RATNER, P.C.

By: Richard A. Levy
Dana Lossia
Robert H. Stroup

80 Eighth Avenue
New York, NY 10011
(212) 627-8100
(212) 627-8182 (fax)
rlevy@lrbpc.com
dlossia@lrbpc.com
rstroup@lrbpc.com

CENTER FOR CONSTITUTIONAL RIGHTS
Darius Charney
666 Broadway, 7th Floor
New York, NY 10012-2399
(212) 614-6438
(212) 614-6499 (fax)
dcharney@ccrjustice.org

SCOTT + SCOTT, LLP
Judy Scolnick
Amanda Lawrence
29 West 57th Street
New York, NY 10019
(212) 223-6444
(212) 223-6334 (fax)
jscolnick@scott-scott.com
alawrence@scott-scott.com

Attorneys for Plaintiffs-Intervenors