

UNITED STATES DISTRICT COURT  
EASTERN DISTRICT OF NEW YORK

-----X  
UNITED STATES OF AMERICA,

Plaintiff,

-and-

THE VULCAN SOCIETY, INC., ET AL.,

Plaintiffs-Intervenors,

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

-against-

ECF Case

THE CITY OF NEW YORK, ET AL.,

Defendants.

-----X

**RULE 56.1 STATEMENT OF UNDISPUTED FACTS**

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, 8<sup>th</sup> 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*

Date of Service: February 2, 2009

## TABLE OF CONTENTS

	<u>Page</u>
Introduction .....	1
Background of the Case .....	2
History .....	2
The Equal Employment Practices Commission’s Audits .....	3
EEOC Charges and Complaint .....	3
Overview of the Administration of Exams 7029 and 2043 .....	5
<b>I. DEFENDANTS’ ADMISSIONS, AS WELL AS UNCONTESTED FACTS AND UNCONTESTED EXPERT OPINIONS, ESTABLISH THAT THE CHALLENGED EXAMS, AS USED, HAD AN ADVERSE IMPACT AGAINST BLACK APPLICANTS .....</b>	<b>9</b>
<b>A. The Disparities In Passing Rates And Eligibility List Rankings Of White And Black Candidates On Exams 7029 And 2043 Were Statistically Significant .....</b>	<b>9</b>
<b>B. The Disparities In Passing Rates And Eligibility List Rankings Of White And Black Candidates On Exams 7029 And 2043 Also Had Practical Significance .....</b>	<b>12</b>
<b>C. Defendants’ “Sample Size” Argument Does Not Eliminate Adverse Impact .....</b>	<b>13</b>
<b>D. Defendants’ Application of The “80 Percent Rule” Does Not Eliminate Adverse Impact .....</b>	<b>14</b>
<b>E. Defendants’ Expert’s Admissions On The Adverse Impact Of Exam 7029 .....</b>	<b>16</b>
<b>II. UNCONTESTED FACTS AND UNCONTESTED EXPERT OPINIONS ESTABLISH THAT EXAMS 7029 AND 2043 WERE NOT JOB RELATED OR CONSISTENT WITH BUSINESS NECESSITY .....</b>	<b>17</b>
<b>A. The Use Of Exams 7029 and 2043 Was Not Validated .....</b>	<b>17</b>
<b>i. The Pass Marks (Cutoff Scores) For Exams 7029 And 2043 Were Never Validated .....</b>	<b>17</b>
<b>ii. The Ranking Of Applicants Was Not Validated .....</b>	<b>21</b>
<b>iii. The Calculation of Applicants Scores Was Flawed .....</b>	<b>23</b>
<b>B. Exams 7029 and 2043 Lacked Content Validity .....</b>	<b>24</b>

i.	The Job Analysis For Exam 7029 Fails To Meet Professional Standards .....	25
ii.	The Abilities Measured Were Not Representative Of Those Needed For the Job .....	30
iii.	The Challenged Exams Were Drafted By Laymen And The Reading Level Of The Questions Was Discriminatorily High .....	35
iv.	Serious Flaws In Test Construction Led To Exams That Did Not Accurately Measure Applicants' Abilities .....	36
C.	Validity Generalization Cannot Be Relied Upon .....	37
III.	DEFENDANTS AND THEIR EXPERTS ADMIT THAT ALTERNATIVE SELECTION PROCEDURES WITH LESS ADVERSE IMPACT AND EQUAL OR GREATER VALIDITY COULD HAVE BEEN USED IN 1999 AND 2002 .....	40
A.	Exam 6019 Had Less Adverse Impact Than Exams 7029 And 2043 .....	40
B.	Exam 6019 Is Equally Or More Valid Than Exams 7029 And 2043 .....	41
C.	Exam 6019 Could Have Been Used In 1999 And 2002 .....	41
D.	Numerous Other Alternative Measures Were Available to Defendants In 1999 And 2002 When The Challenged Exams Were Given .....	42

UNITED STATES DISTRICT COURT  
EASTERN DISTRICT OF NEW YORK

-----X  
UNITED STATES OF AMERICA,

Plaintiff,

-and-

THE VULCAN SOCIETY, INC., ET AL.,

Plaintiffs-Intervenors,

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

ECF Case

-against-

THE CITY OF NEW YORK, ET AL.,

Defendants.

-----X

**RULE 56.1 STATEMENT OF UNDISPUTED FACTS**

**INTRODUCTION**

This statement of uncontested facts is submitted in support of Plaintiffs-Intervenors' motion for summary judgment. This case – and this motion for summary judgment – involves a claim that entry-level firefighter Examinations No. 7029 and 2043, initially administered in 1999 and 2002, discriminate against black applicants. Following some background information (¶¶1-7) and an overview of the administration of the challenged exams (¶¶8-25), the uncontested material facts are grouped to correspond to the three (3) prong test under Title VII of the Civil Rights Act of 1964 (“Title VII”), as amended, 42 U.S.C. § 2000e, *et seq.*, for proving discrimination in the screening and selection of job applicants. They are as follows:

- I. Those establishing that the written portion Defendants' Exams 7029 and 2043 had an adverse impact on black applicants (¶¶26-52).
- II. Those showing that Defendants have not and cannot establish that the exams were job related and consistent with business necessity (i.e., validated) (¶¶53-146).
- III. Those showing that there were alternative selection devices available to Defendants when written Exams 7029 and 2043 were used that had equal or greater validity and less adverse impact on black applicants (¶¶147-167).

## BACKGROUND OF THE CASE

### HISTORY

1. Plaintiff-Intervenor the Vulcan Society acted as plaintiff in a similar action against the FDNY in the 1970s, in which Judge Weinfeld of the Southern District found that the FDNY's written entry-level firefighter examination had a racially discriminatory impact and was not sufficiently job related to withstand attack under 42 U.S.C. § 1983. That decision was affirmed by the Second Circuit Court of Appeals in 1973, which also upheld the District Court's injunction requiring the City to hire one (1) minority candidate for every three (3) white candidates. *See Vulcan Soc. of New York City Fire Dept., Inc. v. Civil Service Com.*, 490 F.2d 387 (2d Cir. 1973). In the early 1970s, New York City's firefighting force was made up of five percent (5%) minority firefighters. *Id.* at 398. After the expiration of the Court's injunction in 1977, however, the hiring of black firefighters plummeted to the point that, as of March 31, 2005 (the most recent date for which such data was produced), black firefighters made up only three percent (3%) of the force. (App. A:1, Demographics of FDNY firefighters).<sup>1</sup> Meanwhile New York City is made up of 26.6% black residents according to the most recent U.S. census data. (App. B:2, U.S. Census data).

2. Hiring rates from the two exams given just *prior* to those challenged here were consistently low. The total number of black hires from Exam 7022 (given in 1988) was 29 out of

---

<sup>1</sup> 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. TT:600". 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."

2,256 (a black hiring rate of less than 1.3%). (App. C:3, appointments from Exams 7022 and 0084 by race). The next exam, Exam 0084 (given in 1992) resulted in 56 black hires out of 2,692 total hires (a black hiring rate of 2%). *Id.*

#### THE EQUAL EMPLOYMENT PRACTICES COMMISSION'S AUDITS

3. The New York City Equal Employment Practices Commission ("EEPC"), a non-mayoral City agency that monitors enforcement of the City's Equal Employment Opportunity Policy, began an audit of firefighter Exam 7029 in the fall of 1999. According to the EEPC's research at that time, which was provided to the FDNY, employment rates of black firefighters were much lower in New York City than in any other major city in America. (App. D:4). As of 1999, the percentage of black firefighters in Chicago was 20%; in Houston was 17%; in Los Angeles was 14%; in Philadelphia was 26%; and in San Diego was 8%. *Id.* The EEPC's Executive Director, Abraham May, testified that the FDNY had – and still has – the worst minority under-representation of any New York City agency he has seen. (App. E:5-8, May Tr. 49-51). The EEPC recommended that the FDNY conduct an adverse impact study on the written portion of Exam 7029. The FDNY, and ultimately Mayor Bloomberg, refused to do so. (App. F:9-16, EEPC Report and Mayor's response).

#### EEOC CHARGES AND COMPLAINT

4. On August 9, 2002, the Vulcan Society filed a charge with the United States Equal Employment Opportunities Commission ("EEOC") alleging that firefighter Exam 7029 had an adverse impact against black applicants and also alleging intentional discrimination. (App. G:17-23, EEOC Charge by Vulcan Society). In the EEOC's Determination of the charge, it concluded that an analysis of test scores on Written Exam 7029 "indicate a high degree of adverse impact against African-American applicants, and all the differences in percentages

between blacks and whites are highly statistically significant.” (App. H:27, EEOC Determination). The EEOC also found that the exam was not validated (i.e., had not been shown to predict job performance). *Id.*

5. On February 24, 2005, three additional charges of discrimination were filed by Roger Gregg, Marcus Haywood and Candido Nuñez, black men who were denied appointment after taking, and passing, Exam 2043. (App. I:30-51, EEOC Charges by Gregg, Haywood and Nuñez). The EEOC’s Determinations of their charges found that Exam 2043, like Exam 7029, had adverse impact on black applicants and was not validated. (App. J, EEOC Determinations).

6. The City refused to engage in conciliation at the EEOC (App. K:62, City letter to EEOC), and the charges filed by the Vulcan Society, Gregg, Haywood and Nuñez were all referred to the United States Department of Justice, leading to the initiation of this lawsuit. The United States’ filed its Complaint in this matter on May 21, 2007. (Dkt. 1). On September 5, 2007, the Court granted Plaintiffs-Intervenors leave to intervene not only with respect to the issues raised in the United States’ Complaint but also with respect to their additional claims of disparate impact and disparate treatment against the City, the FDNY, the Department of Citywide Administrative Services (“DCAS”), Fire Commissioner Nicholas Scoppetta and Mayor Michael Bloomberg. (Dkt 47). Plaintiffs-Intervenors’ Complaint, filed September 25, 2007, alleges, *inter alia*, discrimination in selection procedures in violation of Title VII, as well as 42 U.S.C. § 1981, 42 U.S.C. § 1983, the New York State Human Rights Law, New York Executive Law §§ 290 and 296, and New York City Administrative Code §§ 8-101, *et seq.* (Dkt. 48).

7. Essentially, the Complaint charges Defendants with using firefighter selection procedures that have a statistically significant adverse impact against black applicants, that are not job related or consistent with business necessary, and for which lawful alternatives were

available. The Complaint also alleges that Defendants maintained these procedures despite specific knowledge of their adverse impact on blacks and with reckless disregard as to their unlawfulness under Title VII and other applicable laws.

#### OVERVIEW OF THE ADMINISTRATION OF EXAMS 7029 AND 2043

8. Prior to the administration of a civil service examination, the City issues a Notice of Examination (“NOE”), setting forth the requirements for the position and other details about how to apply to take the test. (App. L:63-73, NOEs for Exams 7029, 2043 and 6019).

9. Those interested in taking Exam 7029 or 2043 were required to return a completed application and application fee to DCAS, unless the fee was waived as set forth by City regulations. (App. M:90, Adm. 48-50). Applicants were then sent admission cards. (App. M:91, Adm. 51-52).

10. Exams 7029 and 2043 each had two components: a written, multiple choice test and a physical performance test (“PPT”). (App. L:64, 69). The written and physical components of each test were intended to be worth 50% of a candidate’s overall score. (App. L:64, 69). As discussed in paragraphs 75-78 below, due to flaws in Defendants’ method of computation, the combined score did not accurately reflect a 50-50 weighting of the written test and the PPT.

11. The written portion of both Exam 7029 and Exam 2043 consisted of a paper and pencil test with 85 multiple choice questions. (App. M:85-87, Adm. 31, 38). Each candidate’s score on Written Exam 7029 and Written Exam 2043 was the percentage of the 85 questions on the examination that the candidate answered correctly. (App. M:92, Adm. 58).

12. Written Exams 7029 and 2043 were based on the same job analysis and test plan (App. CC:393, Adm. 11; App. KK:537-543, Johnston Tr. 19, 23-28), and Defendants’ expert Dr. Philip Bobko testified that he considered them to be “very similar,” based both upon his own



review as well as a representation by Defendants. (App. T:225-227, Bobko Tr. 296-298).

13. Nevertheless, the passing score on Written Exam 7029 was 84.705 (App. M:94, Adm. 65), while the passing score on Written Exam 2043 was 70.000. (App. N:109, Interrog. 36).

14. The City did not allow candidates who scored below 84.705 on Written Exam 7029 or below 70.000 on Written Exam 2043 to take the PPT. (App. M:95, Adm. 68-69).

15. The written portion of Exam 7029 was first given to applicants on February 26, 1999. (App. N:109, Interrog. 36). Written Exam 7029 was subsequently administered to additional groups of applicants as late as December 2002. (App. M:91-92, Adm. 55). Candidates were hired from the Exam 7029 eligibility list at least until December 2004. (Answer, Dkt. 8 at ¶10; App. O:119, Siskin I at 12).<sup>2</sup>

16. The written portion of Exam 2043 was first given to applicants on December 14, 2002. (App. M:92, Adm. 56). The City continued to give Written Exam 2043 to other groups of applicants at least until March 2007. (App. M:92, Adm. 57). Candidates were hired from the Exam 2043 eligibility list at least until August 5, 2007. (App. O:137, Siskin I at Table 12A; App. P:148, Dep. of Thomas Patitucci, DCAS Asst. Commissioner for Examinations (designee),

---

<sup>2</sup> Dr. Bernard R. Siskin, Ph.D. was retained as an expert by the United States in this action. Dr. Siskin's November 2007 expert report on adverse impact is referred to here as "Siskin I at \_\_\_," and his July 2008 expert report on job relatedness and business necessity is referred to as "Siskin II at \_\_\_." The United States' additional experts are David P. Jones, Ph.D. and Leaetta M. Hough, Ph.D., whose July 2008 expert report on job relatedness and business necessity is referred to as "Jones-Hough at \_\_\_," and Irwin L. Goldstein, Ph.D., whose July 2008 expert report on content validity is referred to as "Goldstein at \_\_\_." Joel P. Wiesen, PhD. was retained as an expert by Plaintiffs-Intervenors. Dr. Wiesen's January 2008 expert report on adverse impact is referred to here as "Wiesen I at \_\_\_," and his August 2008 expert report on job relatedness and business necessity is referred to as "Wiesen II at \_\_\_." Drs. Philip Bobko and F. Mark Schemmer were retained as expert witnesses in this action by Defendants. Their January 2008 expert report is referred to here as "Bobko-Schemmer at \_\_\_".

October 9, 2007, Tr. 218:4-9).

17. Before inviting candidates who passed Written Exam 7029 or 2043 to take the PPT, the City sent them each a Notice of Result, informing them of their scores on the written test and whether they had passed. (App. M:92-93, Adm. 59-60).

18. After Notices of Result were sent, there was an appeals period during which a candidate could challenge the accuracy of his/her score on the written examination. (App. M:93-94, Adm. 61-63). If a candidate's written test score was corrected as a result of an appeal, the candidate would be sent a new Notice of Result, reflecting the corrected written test score. (App. M:94, Adm. 64).

19. Each candidate who passed Written Exam 7029 or 2043 was sent a packet with instructions regarding the PPT. (App. N:109, Interrog. 36). The City used the same PPT for both Exam 7029 and Exam 2043. (App. M:88-89, Adm. 40, 43). The PPT consisted of eight (8) physical tasks or "events." (App. M:88, Adm. 41). A candidate had to pass any six (6) of the eight (8) events in order to pass the PPT. (App. M:89, Adm. 44). A candidate's score on the PPT was the percentage of the PPT events he or she passed. (App. M:89, Adm. 45). There were only three (3) possible passing scores on the PPT. Those who passed all eight (8) events scored 100%; those who passed seven (7) out of the eight (8) events scored 87.5%; those who passed six (6) out of the eight (8) events scored 75%. (App. M:89, Adm. 46). Candidates who failed more than two (2) events were not permitted to continue the PPT and were given a score of 62.5%. (App. M:90, Adm. 47).

20. Candidates' scores on the Written Exam and the PPT were combined to arrive at a "transformed score" and certain additional points were then added for Veterans Credit, Residency Credit and/or Legacy Credit to arrive at the "Adjusted Final Average score." (App.

M:97-98, Adm. 86, 96).

21. Candidates who passed both the written and PPT components of Exam 7029 or 2043 were sent an updated Notice of Result informing them of their Adjusted Final Average and their rank on the eligibility list, but not informing them of their likelihood of being reached for appointment. (App. I:47-48, 50-51).

22. The Exam 7029 eligibility list was established on November 15, 2000. (App. N:32, Interrog. 36). The Exam 2043 eligibility list was established on May 5, 2004. (App. M:95-96, Adm. 71).

23. Based on the number of firefighters needed to fill the FDNY's next academy class, DCAS would "certify" a group of candidates from the eligibility list, beginning with the highest ranked candidate on the list. (App. Q:150-151, Dep. of Stephen Dobrowsky, Deputy Asst. Commissioner for Civil Service Administration, Tr. 59-60; App. M:99-101, Adm. 105, 110-111).

24. These "certified" candidates were then investigated by the FDNY's Candidate Investigation Division ("CID") for potential hire. The CID determined whether the candidates on a "certification list" met the requirements for appointment listed on the NOE, including checking criminal background, education or military service, employment references, Certified First Responder with Defibrillation ("CFR-D") certification, English-speaking ability, citizenship, age, proof of identification and driving record. (App. M:98, Adm. 98; App. N:111, Interrog. 36).

25. The City appointed candidates in descending rank order (starting with list number "1") from among those on the certification list who had completed the investigation process, had been determined to be qualified at the time a new academy class was appointed, and accepted an

offer of employment. (App. M:101, Adm. 112). If a candidate's list number was not reached by the time of the last appointment to a given academy class, the candidate would not be appointed at that time, even if s/he had completed all steps in the selection process and been found qualified. (App. M:101, Adm. 113).

**I. DEFENDANTS' ADMISSIONS, AS WELL AS UNCONTESTED FACTS AND UNCONTESTED EXPERT OPINIONS, ESTABLISH THAT THE CHALLENGED EXAMS, AS USED, HAD AN ADVERSE IMPACT AGAINST BLACK APPLICANTS**

**A. The Disparities In Passing Rates and Eligibility List Rankings of White And Black Candidates on Exams 7029 and 2043 Were Statistically Significant**

26. In the fields of statistics, test development, and other social sciences, it is generally accepted that a disparity in outcomes between two subgroups is "statistically significant" if there is a five percent (5%) or lower likelihood that such disparity could have occurred by chance. (App. O:120, Siskin I at 15; App. 152, Wiesen I at 12; App. S:167, Bobko-Schemmer at 11). This "5% likelihood" threshold corresponds to roughly 1.96 units of standard deviation, while 2.56 units of standard deviation correspond to a one percent (1%) likelihood that a disparity is due to chance. *Id.* Thus, social scientists and the federal courts generally consider disparities of at least two (2) or three (3) standard deviations to be statistically significant. *Id.*<sup>3</sup>

**Exam 7029**

27. Approximately 12,915 white candidates and 1,749 black candidates took the written portion of Exam 7029. (App. O:132, Siskin I at Table 1; App. R:154, Wiesen I at 18).

---

<sup>3</sup> See also Castaneda v. Partida, 430 U.S. 482, 496-97 n.17 (1977).

28. Of the white candidates who took the written portion of Exam 7029, 89.9% passed, while only 60.3% of black candidates who took the written portion of Exam 7029 passed. (App. O:132, Siskin I at Table 1; App. R:154, Wiesen I at 18).

29. Defendants' experts did not contest, in either their report or at deposition, the figures or calculations presented in paragraphs 27 and 28 above.

30. Plaintiffs-Intervenors' and Plaintiff's experts showed – and Defendants admit – that the disparity between the pass rate of black and white candidates on the written portion of Exam 7029 exceeded three (3) units of standard deviation. (App. M:74, Adm. 1; App. S:170-171, Bobko-Schemmer at 17-18; *see also* App. O:115, 121-122, 132, Siskin I at 3, 21-22, Table 1; App. R:154-155, Wiesen I at 18-19).<sup>4</sup>

31. Defendants also admit that as a group, black candidates who passed both the written and physical portions of Exam 7029 were ranked lower on the eligibility list for Exam 7029 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 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).<sup>5</sup>

### **Exam 2043**

32. Approximately 13,878 white candidates and 1,393 black candidates took the written portion of Exam 2043. (App. O:135, Siskin I at Table 5; App. R:158, Wiesen I at 42).

---

<sup>4</sup> Plaintiff's and Plaintiffs-Intervenors' experts actually calculated the disparity to be more than thirty-three (33) units of standard deviation. (App. O:121, 132, Siskin I at 21, Table 1; App. R:155, Wiesen I at 19). This means that the odds are much less than 1 in 10,000 that the disparity would occur by chance. (App. R:155, Wiesen I at 19).

<sup>5</sup> Plaintiff's and Plaintiffs-Intervenors' experts actually calculated the disparity to be more than six (6) units of standard deviation. (App. O:124, Siskin I at 24; App. R:156, Wiesen I at 27).

33. Of the white candidates who took the written portion of Exam 20439, 7.2% passed, while only 85.4% of black candidates who took the written portion of Exam 2043 passed. (App. O:135, Siskin I at Table 5; App R:158, Wiesen I at 42).

34. Defendants' experts did not contest, in either their report or at deposition, the figures or calculations presented in paragraphs 32 and 33 above.

35. Plaintiffs-Intervenors' and Plaintiff's experts proved – and Defendants admit – that the disparity between the pass rates of white and black candidates on the written portion of Exam 2043 exceeded three (3) units of standard deviation. (App. M:75, Adm. 3; App. S:170-171, Bobko-Schemmer at 17-18; *see also* App. O:117, 126, 135, Siskin I at 5, 26, Table 5; App. R:158-159, Wiesen I at 42-43).<sup>6</sup>

36. Plaintiff's, Plaintiffs-Intervenors' and Defendants' experts all concluded that, as a group, black candidates who passed both the written and physical portions of Exam 2043 were ranked lower on the eligibility list for Exam 2043 (i.e., had higher list numbers) 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 exceeded 2.9 units of standard deviation. (App. 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).<sup>7</sup>

37. Plaintiff's and Defendants' experts both concluded that the disparity in the proportions of black and white candidates on the Exam 2043 eligibility list ranked at or above the

---

<sup>6</sup> Plaintiff's and Plaintiffs-Intervenors' experts actually calculated the disparity to be more than twenty-one (21) units of standard deviation. (App. O:126, Siskin I at 26; App. R:159, Wiesen I at 43).

<sup>7</sup> Plaintiff's and Plaintiffs-Intervenors' experts actually calculated the disparity to be more than nine (9) units of standard deviation. (App. O:128-129, Siskin I at 31-32; App. R:160, Wiesen I at 51).

rank of 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).<sup>8</sup>

**B. The Disparities In Passing Rates and Eligibility List Rankings of White And Black Candidates on Exams 7029 and 2043 Also Had Practical Significance**

38. Defendants' use of the written portion of Exam 7029 as a pass/fail screening device with a cutoff score of 84.705 also resulted in a shortfall of about 457 black test passers. In other words, approximately 457 additional black candidates who took the written portion of Exam 7029 would have passed if black candidates had passed the exam at the same rate as white candidates. (App. O:122, Siskin I at 22; App.R:155, Wiesen I at 19; App. T:197-200, Bobko Tr. 139, 142-143, 145).

39. Defendants' use of the written portion of Exam 2043 as a pass/fail screening device with a cutoff score of 70.000 resulted in a shortfall of about 150 black test passers. In other words, approximately 150 additional black candidates who took the written portion of Exam 2043 would have passed if black candidates passed the exam at the same rate as white candidates. (App. O:127, Siskin I at 27; App. R:159, Wiesen I at 43; App. T:197-200, Bobko Tr. 139, 142-143, 145).

40. As a result of Defendants' rank-order processing and selection of firefighter candidates from the eligibility list for Exam 7029, approximately 68 black candidates who passed both the written and physical portions of Exam 7029 were delayed in being hired as firefighters. (App. O:116, 125, 133-134, Siskin I at 4, 25, Tables 3A and 3B; App. T:201-204, Bobko Tr. 153-156). These 68 firefighters lost approximately 20 years of combined employment, or about 3 ½ months of wages, benefits and seniority each. (App. O:116, Siskin I at 4).

---

<sup>8</sup> Plaintiff's expert actually calculated the disparity to be more than nine (9) units of standard deviation. (App. O:129-130, Siskin I at 32-33).

41. Defendants' rank-order processing and selection of firefighter candidates from the eligibility list for Exam 2043 resulted in a hiring shortfall of approximately 30 to 40 black candidates. In other words, approximately 30 to 40 additional black candidates from the 2043 eligibility list would have been hired as firefighters if the proportion of black candidates on the eligibility list ranked at or above the rank of the last person appointed from that list was the same as the proportion of white candidates on the eligibility list ranked at or above the rank of the last person appointed from that list. (App. O:129-130, 136, Siskin I at 32-33, Table 11; App. T:197-200, Bobko Tr. 139, 142-143, 145).

42. As a result of Defendants' rank-order processing and selection of firefighter candidates from the eligibility list for Exam 2043, approximately 44 black candidates who passed both the written and physical portions of Exam 2043 were delayed in being hired as firefighters. (App. O:130-131, 138, Siskin I at 33-34, Table 12B; App. T:201-204, Bobko Tr. 153-156). For example, Candido Nuñez, a Plaintiff-Intervenor in this matter, was ranked number 5,003 on the Exam 2043 eligibility list (App. I:47) and was only reached for hire in the very last class hired from that list in January 2008, more than five (5) years after he took Written Exam 2043 in December 2002. (App. TT:599-600, Nuñez Tr. 58-59).

**C. Defendants' "Sample Size" Argument Does Not Eliminate Adverse Impact**

43. Defendants' experts contend that the very high level of statistical significance of the disparities between the pass rates and eligibility list rankings of black and white candidates on each of the challenged exams (see footnotes 4-8 supra) was artificially inflated by the large sample sizes, i.e., the large numbers of white and black applicants for each of the two exams. (App. S:168-169, 177-179, Bobko-Schemmer at 14-15, Appendix A).



44. However, Defendants' experts conceded that even if the sizes of the relevant samples were reduced by ninety percent (90%), the following would nonetheless be true:

a. the disparity between the pass rates of white and black firefighter candidates on the written portion of Exam 7029 would still exceed three (3) units of standard deviation. (App. O:122, Siskin I at 22; App. S:171, Bobko-Schemmer at 18);

b. the disparity between the pass rates of white and black candidates on the written portion of Exam 2043 would still exceed three (3) units of standard deviation. (App. O:126-127, Siskin I at 26-27; App. S:170-171, Bobko-Schemmer at 17-18);

c. the disparity between the average rank of white candidates and the average rank of black candidates on the eligibility list for Exam 7029 would still exceed two (2) units of standard deviation. (App. S:170-171, Bobko-Schemmer at 17-18);

d. the disparity between the average rank of white candidates and the average rank of black candidates on the eligibility list for Exam 2043 would still exceed 2.9 units of standard deviation. (App. S:170, 172, Bobko-Schemmer at 17, 19); and

e. the disparity between the proportions of black and white candidates on the Exam 2043 eligibility list ranked at or above the rank of the last person appointed from that list would still exceed three (3) units of standard deviation. (App. S:171, Bobko-Schemmer at 18).

**D. Defendants' Application of The "80 Percent Rule" Does Not Eliminate Adverse Impact**

45. Defendants' expert report places great emphasis on the fact that some of the disparities between the pass rates and eligibility list rankings of the relevant demographic subgroups of applicants on the two challenged exams did not violate the so-called "80%" or "four-fifths" rule of thumb for assessing adverse impact. (App. S:164-167, Bobko-Schemmer at 8-11). Under

this rule of thumb, a selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) or 80% of the rate for the group with the highest rate “will generally be regarded” as evidence of adverse impact. See *Uniform Guidelines on Employee Selection Procedures* (“*Uniform Guidelines*”) 29 C.F.R. § 1607.4D.

46. However, Defendants’ expert Dr. Bobko has acknowledged both in his academic writing and during his deposition that when dealing with large samples, statistical significance testing is a superior method for controlling for “false negatives,” i.e., incorrect findings of no adverse impact, than is the 80% rule. (App. T:191-194, Bobko Tr. 61-62, 65-66).<sup>9</sup>

47. More importantly, Defendants’ use of Exams 7029 and 2043 in the New York City firefighter hiring process did in fact violate the 80% Rule, and they admit that much, as follows:

a. The ratio of the pass rate of black candidates to the pass rate of white candidates on the written portion of Exam 7029 was less than 80%. (App. M:80, Adm. 13);

b. The proportion of black candidates on the eligibility list for Exam 2043 ranked at or above the rank of the last person appointed from that list was less than 80% of the proportion of white candidates on the eligibility list for Exam 2043 who ranked at or above the rank of the last person appointed from that list. (App. M:81-82, Adm. 16-17; App. O:136, Siskin I at Table 11; App. R:162, Wiesen I at 53).

---

<sup>9</sup> In addition, both the Second Circuit Court and the federal agencies in charge of enforcing Title VII have concluded that even in cases where an employment test does not fail the “80% Rule,” its adverse impact can still be established if the disparities in outcomes between the relevant sub-groups are statistically and practically significant. See *Uniform Guidelines*, 29 C.F.R. §1607.4D; *Smith v. Xerox Corp.*, 196 F.3d 358, 365 (2d Cir. 1999).

**E. Defendants' Expert's Admissions on the Adverse Impact of Exam 7029**

48. Defendants' expert Dr. Bobko, admitted during his deposition that Defendants' use of the written portion of Exam 7029 as a pass/fail screening device had an adverse impact on black firefighter candidates. (App. T:185, Bobko Tr. 52).

49. Dr. Bobko also admitted during his deposition that Defendants' rank-order processing and selection of candidates from the eligibility list for Exam 7029 had an adverse impact on black firefighter candidates. (App. T:189, Bobko Tr. 56).

50. While both Plaintiff's and Plaintiffs-Intervenors' experts concluded that Defendants' use of the written portion of Exam 2043 as a pass/fail screening device had an adverse impact on black firefighter candidates (App. O:118, Siskin I at 10; App. R:159, Wiesen I at 43), Defendants' expert Dr. Bobko declined to offer any opinion on this question. (App. T:187-188, Bobko Tr. 54-55).<sup>10</sup>

51. While both Plaintiff's and Plaintiffs-Intervenors' experts concluded that Defendants' rank-order processing and selection of candidates from the eligibility list for Exam 2043 had an adverse impact on black firefighter candidates (App. O:118, Siskin I at 10; App. R:161, Wiesen I at 52), Dr. Bobko declined to offer any opinion on this question. (App. T:190, Bobko Tr. 57)<sup>11</sup>.

---

<sup>10</sup> When asked at deposition for his opinion on this question, Dr. Bobko refused to answer, stating that he needed "additional clarification" on the definition of "adverse impact." (App. T:187-188, Bobko Tr. 54-55). However, when asked only a few moments earlier whether, in his opinion, Defendants use of the written portion of Exam 7029 as a pass/fail screening device had an adverse impact on black candidates, Dr. Bobko answered "yes" without seeking any additional clarification on the meaning of "adverse impact." (App. T:185, Bobko Tr. 52).

<sup>11</sup> Again, when asked for his opinion on this question, Dr. Bobko refused to answer and stated that he needed clarification on the definition of adverse impact. (App. T:190, Bobko Tr. 57).

52. Nowhere in Defendants' experts' report do Drs. Bobko or Schemmer *deny* that the use of Exams 7029 and 2043 as pass/fail screening devices and in rank ordering candidates with passing scores had an adverse impact on black candidates.

**II. UNCONTESTED FACTS AND UNCONTESTED EXPERT OPINIONS ESTABLISH THAT EXAMS 7029 AND 2043 WERE NOT JOB RELATED OR CONSISTENT WITH BUSINESS NECESSITY**

**A. The Use Of Exams 7029 and 2043 Was Not Validated**

53. The validity of an examination is determined not only by its content but also by the way it is used to screen and rank exam takers. (App. U:270, Jones-Hough at 53). If the ranking of applicants on an eligibility list, or the passing score of an examination, is not valid, the examination itself cannot be valid. *Guardians Ass'n. v. Civil Serv. Comm.*, 630 F.2d 79 (2d Cir. 1980).

54. Defendants' expert Dr. Schemmer testified that "validity is as much a function of the process and application of the test as is the test itself, that certainly validity can be affected by an improper application or inference as validity doesn't follow the test around." (App. AA:344, Schemmer Tr. 244).

**i. The Pass Marks (Cutoff Scores) For Exams 7029 and 2043 Were Never Validated**

55. Where an employer has not demonstrated that a one- or two-point difference in scores on an exam reflects a meaningful difference in performance on the job, and where there is adverse impact in passing the exam, the employer cannot lawfully set a cutoff score solely on the basis of the expected number of applicants needed to fill vacancies. *Guardians*, 633 F.2d at 105.

56. Defendants' expert Dr. Schemmer concedes that the passing score of an exam may defeat the exam's job relatedness:

Q. Do you agree with this statement, no matter how valid the exam, it is the cutoff score that ultimately determines whether a person passes or fails. A cutoff score unrelated to job performance may well lead to the rejection of applicants who are fully capable of performing the job.

A. [Dr. Schemmer:] As stated, I would agree with that.

(App. AA:344, Schemmer Tr. 244).

57. Defendants' expert report makes no attempt to justify the pass/fail cutoff scores used on Written Exam 7029 or 2043:

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).

58. Dr. Bobko acknowledged this in his testimony regarding his and Dr. Schemmer's expert report:

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).

59. Nor do Defendants contend that the cutoff scores used on Written Exam 7029 and

Written Exam 2043 measure the minimum level of the tested skills, abilities or other characteristics necessary for successful performance of the job of entry-level firefighter in the FDNY. (App. N:107, Interrog. 30).

60. Carol Wachter, DCAS's Assistant Commissioner for Examinations at the time that Exam 7029 was developed, testified that no validity studies were conducted with respect to the pass mark for Written Exam 7029. (App. W:299, Wachter Tr. 85). 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).

61. Thomas Patitucci also testified that the Written Exam 7029 cutoff score was not based on data gathered 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," and he objected to the fact that the FDNY rather than DCAS was setting the pass mark for the exam. (App. P:141-143, Patitucci Tr. 42-44).

62. The Examiner for Exam 7029 also never conducted an analysis of whether the cutoff score of 84.705 used on Written Exam 7029 corresponds to the level of the tested abilities necessary to perform the firefighter job successfully, nor did he do any analysis of whether a score of 70 on written Exam 2043 corresponds to the minimum level of the tested abilities necessary to perform the firefighter job successfully. (App. X:309-310, Morrongiello Tr. 96-97).

63. The cut-off score for Exam 2043 was not set based on the level of competence necessary for the job, or based on the FDNY's hiring needs, but rather pursuant to DCAS Personnel Rule and Regulation 4.4.9, which sets 70% as a default passing score unless the Notice of Exam of the DCAS Commissioner provide otherwise. (App. N:105-106, Interrog. 16; App. P:140, Patitucci Tr. 20).

64. Dr. Bobko testified that it was represented to him that Exams 7029 and 2043 were "very similar in their content," and his "review of those exams seemed to confirm that they were very similar." (App. T:225-227, Bobko Tr. 296-298). Nevertheless, the pass mark on Exam 7029 was 84.705, while the pass mark on Exam 2043 was 70. (App. M:94, Adm. 65; App. N:109, Interrog. 36). Defendants cannot show – and have not attempted to show – job relatedness and business necessity in the setting of these cutoff scores.

65. 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 PPT, she would have 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, she would have a transformed score of 83.652. *Id.* Applicant B would have a higher transformed score than Applicant A and should be placed higher on the eligibility list. The problem is that Applicant B's score on the written test would have disqualified her from taking the PPT (App. M:95, Adm. 68), even though the majority of applicants who took the PPT scored 100 (App. Z:341, Wiesen II at 82). 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 is not only nonsensical, it also gives unwarranted weight to the discriminatory written test, in spite of Defendants' intention to weigh the written and physical components of the exam equally.

ii. **The Ranking of Applicants Was Not Validated**

66. The City has presented no facts supporting the validity of its rank-order processing of candidates for appointment. Defendants' expert report does not touch upon the question of rank-ordering, and neither of Defendants experts opined as to the validity of rank-ordering in their deposition testimony.

67. When asked directly whether his and Dr. Schemmer's report in this case established the validity of the rank-order processing and selection of candidates on the challenged exams, 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).

68. For a ranking system to be psychometrically valid, it must meaningfully distinguish between candidates – meaning that one- or two-point differences in scores must actually correspond to differences in job performance. *Guardians*, 630 F.2d at 100-101. Here, differences in scores are likely to be due to measurement error or to the unreliability of the



exams, rather than any actual difference in candidates' abilities. (App. V:287-291, Siskin II at 19-23).

69. 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 on the written exams. *Id.* This means that, on both Exam 7029 and Exam 2043, four (4) wrong answers are within the range of normal variation in scores due to chance and may not represent any true difference between individuals. *Id.* But, four (4) wrong answers on Written Exam 7029 would drop a candidate's maximum possible rank on the eligibility list from 1 to 2, 454. *Id.* Four wrong answers on Written Exam 2043 would drop a candidate's maximum rank on the eligibility list from 1 to 1,713. *Id.* Thus, candidates are ranked based on differences that have no psychometric validity.

70. Aside from measurement errors, the reliability of the exams also impacts their validity. Reliability refers to the extent to which an exam will yield the same results when administered more than once to the same applicants. *Guardians*, 630 F.2d at 101-102. Thomas Patitucci, the DCAS Assistant Commissioner for Examinations, testified that if an exam is not reliable it cannot be valid. (App. P:147, Patitucci Tr. 145). Yet Defendants conducted no analysis of the reliability of Exams 7029 and 2043 and presented no evidence to indicate reliability.

71. Many applicants sat for more than one entry-level firefighter exam. Out of the 17,145 applicants who took Written Exam 7029, 2,667 also took Written Exam 2043 and 479 also took Written Exam 6019 (administered by the City in 2007). (App. U:259, Jones-Hough at 39). Out of the 17,817 applicants who took Written Exam 2043, 2,240 also took Written Exam 6019. *Id.* This provides an opportunity to analyze the reliability of the examinations.

72. Among those who took both Written Exam 7029 and Written Exam 2043, 54.8% of those who failed Written Exam 2043 (i.e., scored below 70) nevertheless scored a 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 a 70 or above on Written Exam 2043. Id. Thus, of those who scored below 70 on either exam, 81.3% scored at or above 70 on the other. Id.

73. Likewise, out of the one-hundred and fifty-six (156) candidates who scored below 70 on either Written Exam 7029 or Written Exam 2043 and also took Written Exam 6019, 84.6% passed Written Exam 6019. (App. V:287-288, Siskin II at 19-20).

74. The City offered no expert or other evidence to rebut or explain the lack of reliability described in paragraphs 72 and 73 above.

**iii. The Calculation of Applicants Scores Was Flawed**

75. Defendants' expert report did not address Defendants' method of combining candidates' written and PPT scores. (App. T:232, Bobko Tr. 329:14-19). Yet Dr. Schemmer agreed that when composite scores are used, the supporting evidence must pertain directly to the specific score or score combination used. (App. AA:344-347, Schemmer Tr. 244-247).

76. While the City intended to weigh the written and PPT components of Exams 7029 and 2043 equally (App. L), in neither case did they succeed in doing so. The analysis that establishes this fact is set forth at (App. Z:341-342, Wiesen II at 82-83; App. V:288-290, Siskin II at 20-22). The analysis is not separately set forth here because it has not been contested by Defendants' experts.

77. In fact, Dr. Catherine Cline,<sup>12</sup> who developed the latest firefighter Exam 6019 for

---

<sup>12</sup> Dr. Cline holds a Ph.D. in Industrial/Organizational Psychology, Measurement and Evaluation. (App. BB:365, Cline Dep. Tr. 26). She has been an expert witness on job analysis, test devel-

Defendants, testified that the City did not properly standardize the scores on Exams 7029 and 2043. (App. BB:372-375, Cline Tr. 67-70).

78. As a result, extremely small differences in performance on the PPT could produce unjustifiably large differences in candidates' PPT scores, and therefore their overall rankings. (App. U:271-275, Jones-Hough at 56-60; App. V:290-293, Siskin II 22-25).

**B. Exams 7029 and 2043 Lacked Content Validity**

79. Under the *Uniform Guidelines*, there are three (3) professionally acceptable methods for validating an employee selection device: construct validation, criterion-related validation, and content validation. 29 C.F.R. §1607.5A-B.

80. 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.” §1607.16E. Content validity is “[d]emonstrated by data showing that the selection procedure is representative of important aspects of performance on the job.” §1607.16D.

81. Defendants admit that they did not conduct construct validation or criterion-related validation studies with respect to Exam 7029 or Exam 2043. (App. CC:394-395, 397, Adm. 20, 23, 62). Rather, Defendants appear to rely on a strategy of content validity, although their report does not make clear what theory of validation is being contemplated. (App. S:173-

---

opment and adverse impact. (App. BB:366-367, Cline Tr. 28-29). She is a member of the American Psychological Association, and the Society for Industrial and Organizational Psychology, and is a past president of the Metropolitan Association of Applied Psychology. (App. BB:368-369, Cline Tr. 41, 45).

175, Bobko-Schemmer at 27-29).

82. Dr. Bobko, however, testified that he has never been involved in conducting a content validity study, other than having watched Dr. Cline develop Exam 6019, and that he does not consider himself to be an expert in “content validity, per se.” (App. T:221-223, Bobko Tr. 230-31, 236).

83. Moreover, Defendants admit that DCAS Examiner Matthew Morrongiello, who lead the development of Exam 7029, did not conduct a content validity study – or any other kind of validity study – with respect to that exam. (App. CC:397, Adm. 61; App. X:306-307, Morrongiello Tr. 74-75). The Test Development Report Mr. Morrongiello compiled after the completion of Exam 7029 was not intended to be a validity report or anything other than a factual narrative of the steps that were taken to develop Written Exam 7029. (App. X:308, Morrongiello Tr. 94).

i. **The Job Analysis For Exam 7029 Fails To Meet Professional Standards**

84. When using a content validation strategy, it is essential for examiners first to conduct a careful analysis of the job, so that they may develop test questions that measure abilities that are critical to the performance of the job. (App. DD:404-405, Goldstein at 6-7).

85. Prior to developing Exam 7029, Defendants conducted a “job analysis” of the entry-level firefighter position. (App. EE:416-423, Exam 7029 Test Dev. Report).

86. No separate job analysis was conducted for Exam 2043. The Exam 7029 Job Analysis Report and the Exam 7029 Test Plan were also used to develop Exam 2043. (App. CC:393, Adm. 11; App. KK:537-543, Johnston Tr. 19, 23-28).

87. Examiner Morrongiello conducted observations of six (6) incumbent firefighters.

(App. EE:417, Exam 7029 Test Dev. Report). He then assembled a focus group of firefighter incumbents who reviewed (a) the firefighter “task list” that had been developed during the job analysis for the prior exam (given in 1992), and (b) a list of cognitive abilities taken from the Fleischman ability taxonomy. (App. EE:417-418, Exam 7029 Test Dev. Report).<sup>13</sup>

88. Next, DCAS created a Job Analysis Questionnaire (“JAQ”) which asked incumbent firefighters: (a) to rate the importance of 196 tasks (grouped into 21 task clusters) to the performance of their jobs, (b) to rate the frequency with which they performed each task, and (c) to rate the importance of 21 cognitive abilities, taken from the Fleischman taxonomy, to the overall performance of their jobs. (App. EE:417-422, Exam 7029 Test Dev. Report). Results were collected from 192 incumbents. Id.

89. The possible ratings were:

- 4 – Critical to the performance of my job
- 3 – Important to the performance of my job
- 2 – Somewhat important to the performance of my job
- 1 – Not relevant to the performance of my job

(App. EE:422, Exam 7029 Test Dev. Report).

90. JAQ respondents were not given definitions, explanations or examples concerning the meanings of “somewhat important,” “important,” or “critical.” (App. X:314-316, Morrongiello Tr. 226-228). Nor were they asked to indicate the level of proficiency required for each task or ability. (App. EE:422, 447-448, Exam 7029 Test Dev. Report).

91. JAQ respondents were also not asked whether the rated tasks and abilities were necessary or important to have on the *first day* of the job. (App. EE:448, Exam 7029 Test Dev.

---

<sup>13</sup> The Fleishman taxonomy also includes many non-cognitive abilities, but DCAS used none of them in the job analysis for Exams 7029 and 2043. (App. X:312-333, Morrongiello Tr. 129-130).

Report). In order to establish content validity, the City must show that Exams 7029 and 2043 measured only those abilities that are needed on “Day 1” of the job rather than abilities that can be learned on the job or in training. (App. DD:407, Goldstein at 12). Because Defendants’ job analysis did not gather this information, there is *no* evidence that Written Exams 7029 and 2043 measured abilities that were needed on Day 1, as is professionally necessary. (App. DD:407, Goldstein at 12; *Guidelines*, 29 C.F.R. § 1607.14C(1)).

92. Of the 196 tasks and 21 abilities the JAQ respondents were asked to rate, 111 tasks and 18 abilities received average importance ratings of 2.5 or higher, i.e. between “somewhat important” and “important” to the performance of the job. (App. EE:422, Exam 7029 Test Dev. Report).

93. Twelve (12) incumbent firefighters were then selected to participate in a “Linking Panel.” The Linking Panel was given a list of the 18 cognitive abilities that had received average importance ratings of at least 2.5 on the JAQ, and they were asked to rate the importance of each ability to the performance of each of the 21 task clusters that had been rated by the JAQ respondents. (App. EE:422-423, 467-477, Exam 7029 Test Dev. Report). For example, the Linking Panel members had to rate the importance of each ability (such as Written Comprehension and Inductive Reasoning) to the performance of Task Cluster A “Initial Response to Incidents/Driving,” which included tasks such as “Dons personal protective gear (turnout coat, helmet, boots, breathing apparatus)” and “Sets out traffic cones or flares to block traffic lanes or streets.” (App. FF:490, Linking Panel worksheet; App. EE: 455, Exam 7029 Test Dev. Report).

94. The possible ratings were:

3 – Critical to the performance of the task cluster

- 2 – Important to the performance of the task cluster
- 1 – Somewhat important to the performance of the task cluster
- 0 – Not relevant to the performance of the task cluster

(App. EE: 422, Exam 7029 Test Dev. Report).

95. Linking Panel members were not given definitions, explanations or examples regarding the meaning of “somewhat important,” “important,” or “critical.” (App. X:317-318, Morrongiello Tr. 251-252).

96. Defendants’ expert, Dr. Schemmer testified that raters will often “generalize and not distinguish between or amongst specific abilities.” (App. AA:360-361, Schemmer Tr. 308-309). In fact, several members of the linking panel testified at deposition that they did not understand the definitions of the 18 cognitive abilities they rated:

a. Rater 1 testified that she did not understand the definitions of Problem Sensitivity (App. GG:512), Inductive Reasoning (App. GG:513), Visualization (App. GG:514-515), or Time Sharing (App. GG:515). This Rater also disagreed with her own linkage of Written Comprehension to seven task clusters, stating “again, it has nothing to do with the written word to perform these tasks.” (App. GG:516).

b. Rater 3 testified, with respect to the importance of Inductive Reasoning, that “For all I know, I probably didn’t know what it meant, so I gave it a two since I didn’t know what it was quite frankly.” (App. HH:518). Regarding his rating of the importance of Deductive Reasoning, he testified “I don’t believe I really knew what deductive reasoning was.” (App. HH:519). He also testified, with respect to the application of Deductive Reasoning to a different task cluster, “I didn’t have an explanation of what deductive reasoning was, so to answer it I was just guessing.” (App. HH:522-524). When asked about his rating of the importance of Written Expression to the task cluster “Providing Medical Assistance,” he stated, “. . . there’s no written

expression. I guess it would be irrelevant to the situation of medical . . . I was looking at the wrong one. I would change the answer to zero.” (App. HH:520-521).

c. Rater 8 testified that he was not sure of the difference between Written Expression and Written Comprehension. (App. II:528). He also testified that he was not sure of the meaning of Visualization (App. II:526-527).

d. Rater 10 did not understand the difference between Inductive Reasoning and Deductive Reasoning (App. JJ:530-533). In connection with the ability Spatial Orientation, he stated that “This sounds like a – it’s more of a physical thing, like I mean is it saying that it is important that firefighters wear their glasses? Is that what it’s saying?” (App. JJ:534-535).

97. Linking Panel members were supposed to rate the importance of each ability to the actual performance of each task cluster. However, Morrongiello testified that, although they were not supposed to, the Linking Panel members may have been rating Written Comprehension and Written Expression as important to various tasks because reading is involved in learning to perform the task, or because firefighters may need to write about the task after performing it. (App. X:321-324, Morrongiello Tr. 315-318).

98. Defendants did not analyze the reliability of the Linking Panel ratings or do any systematic analysis of whether they were consistent. (App. X:320, Morrongiello Tr. 299). Morrongiello reviewed the ratings sheets to make sure there were no blanks but did not scrutinize the substance of the ratings. (App. X:323-324, Morrongiello Tr. 317-318). For example, he testified that he probably did not notice that for Rater 9 (one of the linking panel members), the large majority of his ratings were rated 3, “Critical.” Id.

99. Morrongiello combined the average importance ratings for each ability from the JAQ and the Linking Panel and used that number to determine how many questions on Written



Exam 7029 should test for each ability. (App. EE:423, Exam 7029 Test Dev. Report). Because the JAQ was completed by 192 firefighters while the Linking Panel had only 12 firefighters, the combined ratings gave far more weight to individual Linking Panel members than it did to JAQ respondents. As a result, the larger group of firefighters had their responses outweighed by the relatively few members of the Linking Panel. (App. Z:336, Wiesen II at 49). [As noted in paragraph 96, the Linking Panel members were often unclear on the meaning of the abilities they were rating and the nature of their task.]

100. Defendants' expert report, in the section on job relatedness and business necessity, *never mentions Written Exam 2043 at all*. Dr. Bobko testified that he assumed the two exams were very much alike based on the representation of counsel for the City. (App. T:225-227, Bobko Tr. 296-298). However, Defendants' expert report does not contain any analysis or finding regarding the similarity of Exams 7029 and 2043, which leaves the matter of the validity of Exam 2043 entirely unaddressed by Defendants. (App. S:173-175, Bobko-Schemmer at 27-29).

101. In the opinion of experts for the Plaintiffs, the flaws in the job analysis were fatal to the validity of the exams. (App. DD:408-410, Goldstein 13-15; App. U:258, Jones-Hough at 36; App. Z:330, Wiesen II at 31). Defendants never presented any supplemental expert opinion or analysis to rebut these very substantially documented opinions by the Plaintiffs-Intervenors' and Plaintiff's experts.

ii. **The Abilities Measured Were Not Representative Of Those Needed For the Job**

102. There are many different kinds of abilities that Written Exams 7029 and 2043 could have measured. However, Defendants chose a very small subset of cognitive abilities to

test, and the ones that were tested were not necessarily the most important to the job. In fact, the test was so unrepresentative of the job that it cannot be shown to be valid. (App. U:240-258, Jones-Hough at 18-36; App. Z:331-335, Wiesen II at 32-36).

103. The City admits that, all other things being equal, an examination that measures more of the knowledges, skills, abilities or characteristics (“KSACs”) that are important for a job is expected to be more valid than an examination that measures fewer of the KSACs that are important for that job. (App. CC:397, Adm. 63).

104. Dr. Bobko agreed that the content validity of an exam will be weakened if it doesn’t measure as many important KSACs as possible. (App. T:230, Bobko Tr. 317).

105. Despite this, of the 18 cognitive abilities that came out of the job analysis, DCAS chose nine (9) to include on Written Exams 7029 and 2043: Written Comprehension, Written Expression, Memorization, Problem Sensitivity, Deductive Reasoning, Inductive Reasoning, Information Ordering, Spatial Orientation and Visualization. (App. M:84-86, Adm. 28, 35; App. EE:423, Exam 7029 Test Dev. Report).

106. DCAS determined that the rest of the cognitive abilities identified as “somewhat important” or “important” could not be tested in a written, multiple-choice format. (App. EE:423, Exam 7029 Test Dev. Report). For this reason, the two (2) abilities that received the *highest* importance ratings from incumbent firefighters – Oral Comprehension and Oral Expression – were *not* tested by Exam 7029 or Exam 2043.

107. Chief Daniel Nigro, who held the highest uniformed position in the FDNY, Chief of Department, testified that Oral Comprehension and Oral Expression are both very important to the successful performance of the firefighter job. (App. LL:551, Nigro Tr. 81).

108. Defendants’ consultant Dr. Cline testified that tests of oral comprehension had

been developed in the early 1970s, and she had experience working on exams in which applicants were asked to listen to oral questions (on a tape) and respond in writing. (App. BB:382-384, Cline Tr. 163-165).

109. Dr. Cline also determined that four (4) of the nine (9) abilities that DCAS eliminated from Written Exams 7029 and 2043 as untestable could be measured either directly (Speed of Closure, Flexibility of Closure and Perceptual Speed) or indirectly (Time Sharing) on a written test. (App. MM:553).

110. Defendants did not determine what percentage of the job would be represented by testing for only those nine cognitive abilities and no non-cognitive abilities (other than physical).

As Dr. Bobko testified:

Q.: In your opinion, of all the abilities that are required by or that are important to the entry level firefighter job, what percentage do those nine abilities represent?

A. [Dr. Bobko]: I don't know.

Q.: Can you give me an estimate?

A. [Dr. Bobko]: No.

(App. T:228-229, Bobko Tr. 311-312).

111. Moreover, these nine cognitive abilities may not even be the cognitive abilities *most* important for the job. The U.S. Office of Personnel Management found in 1980 that the ability most highly correlated with successful performance of the entry-level firefighter job is Mechanical Ability. (App. Z:334, Wiesen II at 35). Defendants' expert Dr. Schemmer agreed that written tests for Mechanical Ability were available when Exams 7029 and 2043 were used. (App. AA:350, Schemmer Tr. 263). Mechanical Ability, however, was neither given to the JAQ respondents to be rated nor tested for by Exams 7029 or 2043. (App. EE:465-466, Exam 7029

Test Dev. Report).

112. While each of the nine abilities tested on Written Exams 7029 and 2043 is a cognitive ability, (App. M:84, 86, Adm. 29, 36), Dr. Cline's job analysis for Exam 6019 found that non-cognitive "personal attributes" are as important to the job as cognitive abilities. (App. NN:559). She also found that the job had not changed since the challenged exams were used. (App. BB:376-378, Cline Tr. 72-74).

113. In response to this finding, Dr. Schemmer testified as follows:

Q. If I told you that the job analysis that was done for exam 6019 found that physical abilities and personal attributes were as important as cognitive skills for the New York firefighter job, and also found that the job had not changed from the early 90's, would you then agree the job analysis the City conducted for exam 7029 overemphasized cognitive skills?

A. [Dr. Schemmer:] I would certainly come a great deal closer to agreeing with that, yes.

(App. AA:348-349, Schemmer Tr. 256-257).

114. Examples of non-cognitive KSACs are: Resistance to stress, Teamwork, Responsibility, Desire to Learn, Honesty, Cleanliness, Medical Interest (firefighters perform emergency medical procedures), Achievement Orientation, Dependability, and Conscientiousness. (App. CC:398, Adm. 64).

115. Each of these non-cognitive KSACs has been found to be important to the job of firefighter. (App. CC:398-399, Adm. 65).

116. In fact, when asked about the importance of Integrity, Work Standards and Coordination, Chief Nigro testified that they were *very* important to one's performance as a firefighter job. (App. LL:548-550, Nigro Tr. 78-80).

117. Dr. Bobko recognized that cognitive abilities and non-cognitive abilities are both

important to the firefighter job:

Q.: What is the relative importance of physical abilities, cognitive abilities and personal attributes to the successful performance of the FDNY entry level firefighter job?

A. [Bobko:] I don't know. I can tell you *they are all important* based on the job analysis [done for Exam 6019], but that's all I can tell you.

(App. T:183-184, Bobko Tr. 45-46) (emphasis added).

118. However, the firefighters who responded to the JAQ did not have the opportunity to rate the importance of any non-cognitive abilities to their job, and no questions were developed to measure them. (App. EE:465-466, Exam 7029 Test Dev. Report).

119. Underscoring the relative importance of cognitive versus non-cognitive abilities is the structure of Exam 6019, in which only 1/3 of a candidate's score on the exam was based on the cognitive ability questions. A candidate who gave wrong answers on *all* of the cognitive ability items on Written Exam 6019 (including all of the Memorization items), could still pass Written Exam 6019 if the candidate answered all of the other items correctly. (App. V:284-285, Siskin II at 16-17). This demonstrates that measures of cognitive abilities were far over-emphasized on Written Exams 7029 and 2043, which tested for nothing else.

120. [Non-cognitive abilities and characteristics have frequently been rated as more important to the firefighter job than the cognitive abilities the City tested for. The importance of these non-cognitive abilities is discussed further in paragraphs 159-163, in the section of this Statement concerning available alternatives.]

121. In the expert opinions of Drs. Jones, Hough, Goldstein and Wiesen, the under-representation of abilities on Exams 7029 and 2043 rendered them fatally flawed. (App. U:244, Jones-Hough at 22, ¶103; App. DD:411-412, Goldstein at 16-17; App. Z:330, Wiesen II at 31).

122. Defendants experts have never rebutted these conclusions and submitted no supplemental report to address them at all. They cannot do so now that discovery is long closed.

iii. **The Challenged Exams Were Drafted By Laymen And The Reading Level Of The Questions Was Discriminatorily High**

123. Examination questions are commonly referred to as “items”, and those who draft the questions are the “item writers.” Four firefighters and one lieutenant were placed on an Item Writing Panel that wrote the questions for Written Exam 7029. (App. EE:423-424, Exam 7029 Test Dev. Report).

124. DCAS then assembled a Review Panel of one lieutenant and four firefighters who reviewed the items over the course of two (2) days. (App. EE:424, Exam 7029 Test Dev. Report). Defendants did not ask the reviewers (also called subject matter experts or “SMEs”) to independently judge whether each item measured the ability it was originally designed to measure. (App. DD:406, Goldstein at 11).

125. Items were then reviewed by DCAS staffers and finalized by Examiners Morrongiello (for Written Exam 7029) and Johnston (for Written Exam 2043).

126. As Dr. Cline testified, “reading level, especially if it is unnecessary, will in fact augment adverse impact.” (App. BB:370-371, Cline Tr. 64-65).

127. Dr. Wiesen conducted a reading level analysis, which found that the average reading grade level of Written Exams 7029 and 2043 was above the 12th grade and “was too high for the job of Firefighter.” (App. Z:337-338, Wiesen II at 58-59). In fact, both tests contained many questions that required college or graduate school level of reading ability. (App. Z:338-339, Wiesen II at 59-60).

128. Defendants did not conduct a reading level analysis of Written Exam 7029 or

2043. (App. X:325-327, Morrongiello Tr. 456-458; App. KK:545-546, Johnston Tr. 488-489; App. T:238, Bobko Tr. 471). Nor does the report of Defendants' experts attempt to establish that the reading level of the exams was job related or consistent with business necessity.

129. Many applicants did not have a fair opportunity to show their abilities on Exams 7029 and 2043 because the overly high reading level prevented them from understanding what was being asked. (App. Z:340, Wiesen II at 61). Defendants offered no documents or testimony to rebut Dr. Wiesen's expert finding.

**iv. Serious Flaws In Test Construction Led To Exams That Did Not Accurately Measure Applicants' Abilities**

130. On a well-constructed test, items that measure the same ability should be more highly correlated (in terms of those test-takers who answer items correctly or incorrectly) than items that measure different abilities. (App. V:277-278, Siskin II at 5-6; App. T:207-208, Bobko Tr. 188-189; App. BB:386-387, Cline Tr. 322, 325).

131. An analysis of Written Exam 7029 shows that, for eight (8) of the nine (9) abilities that the Written Exam intended to measure, the items used to measure those abilities correlate more highly with items intended to measure other abilities than they do with items intended to measure the same ability. (App. V:278, 294, Siskin II at 6, Table 1). For example, the items intended to measure Deductive Reasoning correlated more highly with items intended to measure Written Expression, Written Comprehension and Spatial Orientation than they did with other items intended to measure Deductive Reasoning. Id..

132. For Written Exam 2043, four (4) of the nine (9) abilities that Defendants intended to measure used items that correlated more highly with other abilities than with items intended to measure the same ability. (App. V:278, 295, Siskin II at 6, Table 2).

133. A “factor analysis” is a statistical test that can be used to identify the individual factors that explain observed correlations between test items. (App. V:279, Siskin II at 7). For an exam that is intended to measure nine (9) abilities, a factor analysis should show that the test items group together into nine (9) or ten (10) factors. (In addition to nine factors corresponding to the nine abilities being measured, a single common factor may also be expected to measure general cognitive ability, i.e., general intelligence). (App. V:279-280, Siskin II at 7-8).

134. Dr. Siskin conducted a factor analysis that revealed, for both Written Exam 7029 and 2043, a large primary factor and a smaller secondary factor, rather than nine (9) or ten (10) distinct factors or ability domains. (App. V:280-282, Siskin II at 8-10). This analysis shows that the questions on Exams 7029 and 2043 did not actually measure the abilities they purported to measure, (App. V:281, Siskin II at 9), further eliminating any claim to content validity.

135. Defendants did not conduct a factor analysis with respect to either Written Exam 7029 or Written Exam 2043. (App. T:207-208, Bobko Tr. 188-189). They offered no evidence to rebut Dr. Siskin’s findings that the Exam questions simply did not measure what they were designed to measure.

136. In addition to the flaws discussed above, Dr. Wiesen has demonstrated that black test-takers on both exams left more questions blank near the end of the test (in the last ten (10) questions than did whites. (App. Z:342-A, Wiesen II at 74). The difference in the number of questions left blank was statistically significant (i.e., not likely to be explained by chance). *Id.* Dr. Wiesen points out that any aspect of the exam that may have contributed to even a one (1) point difference in the scores of blacks versus whites is of great importance. *Id.*

**C. Validity Generalization Cannot Be Relied Upon**

137. Defendants’ expert report cites a theory of “validity generalization” for the



proposition that “cognitive ability tests generally have validity across most jobs.” (App. T:209-210, Bobko Tr. 200-201). Validity generalization, however, is not referred to in the *Uniform Guidelines*. Moreover, as Dr. Bobko put it: “My guess is if we had ten researchers in a room and we asked them to do a validity generalization statement, they would all do something different.” (App. T:209, Bobko Tr. 200).

138. Defendants’ expert report cites only one study, by Barrett, Polomsky and McDaniel (1999), in support of the theory of validity generalization for written cognitive ability examinations. (App. S:173, Bobko-Schemmer at 27; App. T:210-211, 219, Bobko Tr. 201-202; 215; App. SS:591-597, Barrett et al). The Barrett et al study is a meta-analysis of various cognitive ability exams, and it purports to demonstrate that evidence of the validity of one set of cognitive exams should be considered as evidence that other cognitive ability exams have validity. (App. S:173, Bobko-Schemmer at 27).

139. The Barrett meta-analysis studied three types of firefighter exams: cognitive ability tests, mechanical ability tests, and tests combining cognitive and mechanical ability. (App. SS:591-597, Barrett et al). Barrett’s statistical analysis showed that applicable professional standards for finding validity generalization were *not met* for the cognitive ability tests included in his study. (App. U:266-267, Jones-Hough at 49-50; App. SS:594, Barrett et al. at 510).

140. A pitfall of meta-analyses in general has been referred to by Dr. Bobko as “the file drawer phenomenon,” wherein studies that do not produce positive or noteworthy findings are not published and, therefore, do not come to the attention of researchers conducting meta-analyses. (*See, e.g.*, Roth, Bevier, Bobko, Switzer, and Tyler, 2001).

141. While the summary abstract of the Barrett study claims that the meta-analysis was based on more than 100 sample studies, in reality it used only 25. (App. T:213-214, Bobko Tr.

204-205). Nineteen of Barrett's 101 correlation coefficients were drawn from one sample study executed in Columbus, OH (Kriska & Hines, 1984). (App. T:217-218, Bobko Tr. at 210-211). Bobko did not review any of the validity studies that were used in the Barrett meta-analysis, and he does not know how the job performance of those who took the other tests was measured. (App. T:212, Bobko Tr. at 203).

142. While Barrett et al. (1999) describe their sample studies as having been "conducted over the past two decades" (i.e., from 1979 – 1999), the 25 studies reviewed by Barrett et al were actually conducted between 1958 – 1986. (App. T:211-212, 215, Bobko Tr. at 202-203, 207). Thus, none of the studies cited by Barrett et al are less than twenty (20) years old. (App. T:215-216, Bobko Tr. at 207-208).

143. Moreover, the Barrett study left out many validity studies of firefighter selection tests that were conducted prior to the Barrett meta-analysis, including:

- Morris & McDaniel (1989). Report on the Validation Study for the IPMA National Firefighter Examination. Alexandria, VA: International Personnel Management Association.
- Bownas, D. A., Rosse, R. L., & Dunnette, M. D. (1977). Construct Validation of a Selection Battery for the Entry Level Firefighter Position. Personnel Decisions Research Institute Report 15. Washington, D.C.: U.S. Civil Service Commission.
- Heckman, R. W. (1973). Saint Paul Firefighter Validation Study. (Submitted to The City of Saint Paul). Minneapolis, MN: Personnel Decisions, Inc.
- Arvey, R. (1971). Report on Test Validation Study: Minneapolis Civil Service Firefighter Jobs. Minneapolis, MN: Personnel Decisions, Inc.

(App. U:263-264, Jones-Hough at 46-47).

144. All four of these studies reported criterion-related validities that were lower than the average validities reported by Barrett et al. The "comprehensive review" reported by Barrett et al did not incorporate these findings. (App. U:264, Jones-Hough at 47).

145. Aside from Barrett et al, Defendants' expert report mentions validity generalization in connection with the criterion-related validity study conducted in Washington, D.C. by Landy, Jacobs and Associates. At deposition, however, Dr. Bobko testified that he had no opinion as to whether the Landy, Jacobs study from Washington, D.C. was sufficient to establish the validity of Landy's examination for New York, Exam 0084. (App. T:181-182, Bobko Tr. 31-32; App. U:268-269, Jones-Hough at 51-52).

146. While Drs. Bobko and Schemmer make no specific claim, either in their report or in their deposition testimony, that the challenged exams can be supported by the theory of validity generalization, Plaintiff's experts thoroughly demonstrate that validity generalization does not support Exam 7029 or 2043. (App. U:260-269, Jones-Hough at 43-52).

**III. DEFENDANTS AND THEIR EXPERTS ADMIT THAT ALTERNATIVE SELECTION PROCEDURES WITH LESS ADVERSE IMPACT AND EQUAL OR GREATER VALIDITY COULD HAVE BEEN USED IN 1999 AND 2002**

147. Written Exam 6019 was given on January 20, 2007. (App. M:102, Adm. 124). It is a paper and pencil test consisting of 150 "objectively" scored, multiple choice questions. (App. M:103, Adm. 126-127). The Exam 6019 eligibility list was established on June 11, 2008. (App. M:102, Adm. 123).

148. As shown below, Exam 6019 is an example of a firefighter test that could have been used by Defendants in 1999 or 2002 (the years when the challenged exams were given), and would have been at least as valid but with less adverse impact against black applicants.

**A. Exam 6019 Had Less Adverse Impact Than Exams 7029 And 2043**

149. Defendants admit that a greater number of black candidates passed Written Exam 6019 than Written Exams 7029 or 2043, and a greater number of black candidates achieved a higher ranking on Written Exam 6019 than on either Written Exam 7029 or Written Exam 2043.

(App. CC:401-402, Adm. 74-77).

150. Robert Alexander, the DCAS Examiner who worked with Dr. Cline on the development of Exam 6019 (App. OO:561-563, Alexander Tr. 15, 18, 31) also testified that Exam 6019 had less adverse impact on minorities than Exam 7029 or 2043. (App. OO:571, Alexander Tr. 202:7-10). Mr. Alexander's supervisor, William Klimowicz, also reached this conclusion. *Id.*

**B. Exam 6019 Is Equally Or More Valid Than Exams 7029 And 2043**

151. Defendants admit that Written Exam 6019 was at least as job-related and consistent with business necessity as Written Exams 7029 and 2043. (App. CC:400-401, Adm. 70-73). Moreover, Dr. Bobko agreed that Written Exam 6019 met each of the administrative, scoring and security objectives cited in Defendants' expert report as necessary from a business standpoint. (App. T:232-235, Bobko Tr. 329-332).

**C. Exam 6019 Could Have Been Used In 1999 and 2002**

152. Dr. Cline testified that her field – Industrial/Organization Psychology – began taking personality measures seriously as a predictor of job performance in or around 1993. (App. BB:391-392, Cline Tr. 513-514). And, for the past 10–15 years there has been greater use of personality instruments in testing as a way of possibly reducing impact. (App. BB:379-380, Cline Tr. 82-83).

153. Exam 6019 measured the following *non-cognitive* personal attributes: Tolerance for Stress, Adaptability, Tenacity, Integrity, Work Standards, Resilience, Coordination, and Establishing and Maintaining Interpersonal Relationships. (App. L:72).

154. These non-cognitive abilities were tested using situational judgment exercises (“SJEs”). Dr. Cline had used SJEs to test for non-cognitive abilities on an exam for “custodian

engineer,” which she developed for DCAS in 2001 and 2002. (App. BB:363-364, Cline Tr. 14-15).

155. According to Cline, not only do SJE's reduce impact, there has been evidence of their validity for the past ten (10) years:

And, in the last ten years there has been a lot of information that has come out that, in fact, yes situational judgment tests are valid. They do have validity. There has been meta-analyses conducted that shows that the situational judgment exercises do have validity, that they can be extended to other constructs besides judgment. In a stream of research, the personality measures, when added to cognitive measures, also increased the validity while decreasing impact with minority groups.

(App. BB:388-390, Cline Tr. 507-509).

156. Dr. Bobko testified that SJE's were developed by the United States government, as part of a study called Project A, between 1981 and 1991. (App. T:220, Bobko Tr. 225).

**D. Numerous Other Alternative Measures Were Available to Defendants In 1999 and 2002 When The Challenged Exams Were Given**

157. As a general matter, black test-takers will score lower than whites on paper and pencil tests of cognitive ability. (App. T:195-196, Bobko Tr. 121-122).

158. DCAS Examiner Robert Alexander did a literature review which indicated that cognitive tests produced 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). Alexander found examples – dating back to the 1970s and 1980s – of non-cognitive tests that added to the validity of examinations. (App. OO:568-569, Alexander Tr. 353-357).

159. Non-cognitive abilities have often been found to be more important to the firefighter job than cognitive abilities. The U.S. Civil Service Commission published a report in

1977 which listed abilities and characteristics important to the job of firefighter and ranked them based on their importance to the job. (App. PP:577). The report 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. (App. Z:332, Wiesen II at 33). Each of these most important abilities is non-cognitive. Id..

160. In fact, of the twenty (20) abilities and characteristics that the US Civil Service Commission included in its recommended test weighting plan fore firefighters, only four (4) are cognitive: Problem-solving Ability, Mechanical Ability, Verbal Skills (both written and verbal) and Math Skills. (App. Z:332, Wiesen II at 33). Of these, DCAS measured only Problem Solving and Written Comprehension on Exams 7029 and 2043.

161. The Civil Service Commission report listed existing tests (in use prior to 1977) for these non-cognitive abilities for which there was, already at that time, evidence of empirical validity. (App. PP:578-582).

162. Dr. Schemmer agreed that Resistance to Stress, Teamwork, Responsibility, Desire to Learn, Getting along with Others and Honesty, among others, are abilities that are important to the job of firefighter. (App. AA:351, Schemmer Tr. 291). He also agreed that tests for each of these abilities were available in 1999 and 2002 when 7029 and 2043 were administered. Id.

163. O\*Net, a U.S. Department of Labor database of job descriptions, identifies the following abilities as being the ten (10) most important for the job of firefighter: Dependability, Cooperation, Attention to Detail, Self Control, Stress Tolerance, Initiative, Concern for Others, Analytical Thinking, Persistence, and Social Orientation. (App. QQ). Of these, Analytical Thinking is the only cognitive ability, and the only one measured by Written Exams 7029 and 2043.

164. Dr. Cline testified that a predecessor to the O\*Net abilities list has been available since the 1980s. (App. BB:381, Cline Tr. 89).

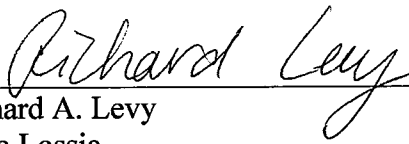
165. Dr. Schemmer reviewed the O\*Net list and agreed that each of the non-cognitive abilities listed in the “top 10” could have been tested for using a paper and pencil test at the time 7029 and 2043 were given and “would certainly have the potential” to add to the validity of the exam. (App. AA:356-358, Schemmer Tr. 299-301). Moreover, “[b]roadly or generally speaking, I would anticipate they would have lesser impact than cognitive assessments.” (App. AA:359, Schemmer Tr. at 302).

166. Dr. Bobko has an advisory relationship with a testing Company called Previsor that develops and markets written tests for the following non-cognitive abilities: Emotional Control, Stress Tolerance, Self Confidence, Adaptability, Independence, Responsibility, Initiative, Competitiveness, Social Awareness, Empathy and Influence. (App. RR; App. T:236-237, Bobko Tr. 459-460).

167. Dr. Schemmer agreed that each of these abilities could have been tested using paper and pencil tests at the time that 7029 and 2043 were given, and he had no reason to believe that the tests for these abilities that Previsor markets are not valid predictors. (App. AA:353-354, Schemmer Tr. 293-94). He testified 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).

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@lrnpc.com  
dlossia@lrnpc.com  
rstroup@lrnpc.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*