

@UNITED STATES DISTRICT COURT
DISTRICT OF MASSACHUSETTS

BRUCE SMITH, PAUL JOSEPH,)
MARTIN JOSEPH, KIM GADDY,)
BRIAN KEITH LATSON, MARWAN)
MOSS, LEIGHTON FACEY, and)
LATEISHA ADAMS,)
)
Plaintiffs,)
)
v.)
)
CITY OF BOSTON, MASSACHUSETTS)
)
Defendant.)

Civil Action No.: 12-10291-WGY

**PLAINTIFFS' POST-TRIAL BRIEF ON EQUALLY VALID BUT LESS
DISCRIMINATORY ALTERNATIVES AND INCORPORATING FINDINGS OF
FACT AND CONCLUSIONS OF LAW**

TABLE OF CONTENTS

I. INTRODUCTION..... 1

II. PLAINTIFFS’ PROPOSED FINDINGS OF FACT REGARDING EQUALLY VALID AND LESS DISCRIMINATORY ALTERNATIVES..... 1

 A. Equally or More Valid Alternatives. 1

 B. Use Of The Multiple Choice Test As A Hurdle/Qualifier, Not As A Strict Rank Order Selection Method..... 3

 C. The Valid Scientific Evidence Demonstrates That The Use Of Alternative Selection Procedures Has Less Adverse Impact On Minority Candidates. 6

 D. Alternative Types Of Selection Procedures Having Less Adverse Impact..... 14

 1. Performance review systems. 14

 2. Structured Oral Interviews 16

 3. Situational Judgment Tests..... 18

III. PROPOSED CONCLUSIONS OF LAW ON LESS DISCRIMINATORY ALTERNATIVES..... 20

 A. The City’s Attempted Anecdotal Evidence That Utilizing Multi-Components Does Not Reduce Adverse Impact Is Not Based Upon Any Valid Evidentiary Showing. 22

 B. The Plaintiffs Have Demonstrated Equally Valid And Less Discriminatory Alternatives That Could Have Been Used. 24

CONCLUSION..... 26

TABLE OF AUTHORITIES

Cases

Allen v. City of Chicago, 351 F.3d 306 (7th Cir. 2003)..... 12, 15, 24, 26

Banos v City of Chicago, 2004 WL 626154 (N.D. Ill. 2004)..... 5

Barnhill v City of Chicago, 2001 WL 243410 (N.D. Ill. 2001)..... 5

Bradley v. City of Lynn, 443 F. Supp. 2d 145 (D.Mass. 2006)21

Bryant v. City of Chicago, 200 F.3d 1092 (7th Cir. 2000) 22

Johnson v. City of Memphis, 770 F.3d 464 (2014) 12, 22, 24

Massachusetts Ass’n of Minority Law Enforcement Officers v. Abban, 434 Mass. 256, 748
N.E.2d 455 (2001)..... 11

I. INTRODUCTION

The Plaintiffs submit their brief now on the issue of “equally valid and less discriminatory alternatives.” The Court granted the Plaintiffs additional time to submit this brief so that the Plaintiffs could conduct discovery on the Defendant’s recently-admitted evidence on the 2014 promotional exam. ECF No. 176. However, to date the Defendant has not produced any of the requested discovery, and therefore this brief is submitted without the benefit of the discovery that this Court granted the Plaintiffs the right to seek. The Defendant’s motion for a protective order on this discovery remains pending with this Court. ECF Nos. 181-82.

II. PLAINTIFFS’ PROPOSED FINDINGS OF FACT REGARDING EQUALLY VALID AND LESS DISCRIMINATORY ALTERNATIVES

A. Equally or More Valid Alternatives.

1. In this case, and in Lopez, there was universal agreement by all experts that adding other components to a written multiple choice job knowledge test significantly increases – and never decreases – validity. Thus, every expert agreed that the adding of job-related components, such as a structured oral interview, or performance review, to the exam process increases validity significantly. (12/17/14, p. 80-83; 12/18/14, p. 89; 12/22/14, p. 16, 106; Lopez Tr. 6:114-15; 16:67; 17:102-105, 111-112, 120-21, 129-30).
2. Dr. Champion testified that he helped develop close to 900 hiring and promotional examinations and used only a written job knowledge test only once, and that was an open book exam for trademark and patent examiners, because he freely admitted that adding components increases validity. (12/22/14, p. 16, 106).

3. Defendant's expert Dr. Silva testified in this trial, and in Lopez, that adding components to an exam typically reduces mean score differences and therefore adverse impact, while increasing validity. (12/18/14, p. 89; Lopez Tr. 17:102-105, 111-112, 120-21, 129-30). His firm, E.B. Jacobs, promotes the use of multi-component tests as leading to "reduced adverse impact" while "increasing validity." (12/18/14, p. 89; Lopez Tr. 17:103-105).
4. Dr. Wiesen and Dr. Hough also testified that adding other components to a multiple choice job knowledge test significantly and dramatically increases validity, particularly since there is relatively low validity, if any, between a multiple choice job knowledge test and performance as a police sergeant or lieutenant.(12/17/14, p. 80-83; 1/6/15, p. 37-38, 67-68). Dr. Outtz testified to the same effect in Lopez. (Lopez Tr. 16:67)
5. The reasons that adding job-related components to a promotional process increases validity, are well documented in the scientific literature. Studies in peer-reviewed journals show that validity increases when an additional component is added, because when one tests for skills and abilities of the job that the written M/C exam cannot assess validity significantly increases. (Ex. 82 at Table 1; 12/17/14, p. 80-81; 12/22/14, p. 105-06; Lopez Tr. 6:114-15 (Dr. Fields); 15:104-106 (Dr. Outtz); 17:111-112 (Dr. Silva)).
6. Indeed, several reputable studies have documented a direct relationship between increased validity and the addition of job-related components, such as the components which plaintiffs proffer in this case. (Ex. 82; Ex. L).
7. Accordingly, on this record, there is simply no dispute that if the Defendants had added additional components to the promotional process for lieutenant, there would have

been a direct proportional increase in validity, and therefore Plaintiffs would satisfy that part of Prong III which requires them to show that alternative procedures would result in an “equally valid” testing process.

8. Additionally, with respect to the multiple choice job knowledge test at issue in this case, the City, Dr. Silva’s firm, E.B. Jacobs, and Dr. Outtz all determined that the multiple choice test should have been “banded” within at least a nine-point spread because there was no way to tell a better-performing candidate. (12/16/14, p. 69; 1/6/15, p. 99; Lopez Tr. 15:59-62; 16:13-14; Lopez Exs. 70-73). Therefore the Defendant must concede that the use of additional components along with at least the banding of the written exam scores rather than strict rank order, was an equally (and likely more valid) valid method of selecting police Lieutenants.

B. Use Of The Multiple Choice Test As A Hurdle/Qualifier, Not As A Strict Rank Order Selection Method.

9. As stated earlier, all of the individuals competing for the position of lieutenant would have already taken and successfully passed (with a reasonably high score) a previous multiple choice job knowledge test, in which 80% of the questions were questions deemed appropriate for the lieutenants’ exam and utilized for that exam. (1/7/15, p. 37; Ex. 8; Lopez Ex. 45).
10. Since all plaintiffs had already been promoted to the position of sergeant, each would have previously scored high enough on that multiple choice exam to have been reached for selection to the position of sergeant. In addition, all have worked as sergeants for at least two years or more and would have learned on the job many of the job knowledges necessary for promotion. (12/15/14, p. 50-51; Ex. 22)

11. In addition, all of the officers who took the lieutenants' exam would have also previously scored relatively high (in the 90's) on the BPD entry-level written cognitive ability test. (1/6/15, p. 29-30).
12. Thus, all of the lieutenant candidates had adequate job knowledge coming into the lieutenants' exam. Moreover, assuming that the pass/fail score was appropriately set, all individuals passing the lieutenants' exam would have demonstrated reasonably proficient job knowledge sufficient to qualify as a police lieutenant.
13. Significantly, no criterion validity study was ever conducted by BPD or its experts to demonstrate that there was any correlation between higher performance on the multiple choice job knowledge test, and performance on the job. Conversely, Dr. Hough testified that in a study she co-authored for the LEAA in the late 1970's, their criterion validity study demonstrated that there was very little correlation between scores on a written job knowledge test and performance as a police officer. (1/6/15, p. 11-15, 28-30). This was confirmed by several studies from reputable journals used to cross examine Dr. Campion. (1/5/15, p. 47-50).
14. Many other cities have used a M/C job knowledge test as a pass/fail device, not as a ranking device, and E.B. Jacobs does this on many occasions, as did Dr. Cassie Fields in her exams, and Dr. Wiesen testified, along with Dr. Outtz and Dr. Campion that this is common. (Lopez Tr. 3:62-63; 6:63-64; 17:93-94; 18:14-15, 62-63). See also Vanguard Justice Soc., Inc. v. Hughes, 592 F. Supp. 245, 248 (D. Md. 1984) (Baltimore); Isabel v. City of Memphis, 404 F.3d 404, 408 (6th Cir. 2005). Indeed for the last twenty years the City of Chicago has successfully used the written M/C test as a qualifying, not ranked component, with each candidate who achieved a qualifying

score being permitted to proceed in the process and then be ranked based on other components, including a merit and assessment component. Barnhill v City of Chicago, 2001 WL 243410 *1 (N.D. Ill. 2001); Banos v City of Chicago, 2004 WL 626154 *1(N.D. Ill. 2004). Indeed, the EEOC Guidelines suggest that such written tests, which are not well documented to show validity, should be used in this fashion. 29 C.F.R. § 1607.5, 1607.14.

15. In addition, Dr. Outtz, Dr. Silva, as well as Dr. Wiesen, testified that they frequently use written multiple choice job knowledge tests as a qualifier for promotional exam, not as a scored device. (12/16/14, p. 86-88; 12/19/14, p. 119; Lopez Tr. 18:13-15)).
16. Further, as described above, since all of the individuals taking the Lieutenants' exam have already shown proficiency in a job knowledge test when becoming Sergeant, and 80% of the questions were the same as on the Lieutenants' exam, it would be reasonable to only use the written multiple choice test, that is known to have such high adverse impact, as a qualifier to establish the pool of candidates.
17. After establishing the pool, selections could be made utilizing a combination of performance review, structured oral interview, and other type assessments. The scientific literature establishes that it is almost certain that this will reduce adverse impact while not impairing validity at all, because by adding these additional components, the test will be more job-related than only one component (i.e. the written test), and since the written test is a non-scored component, any small decrease in validity because multiple choice test results are not being used in rank order fashion, will be far outweighed by the increase in validity by testing for other attributes important to a police lieutenant, but not testable by a written exam.

18. Moreover, the practice of banding is similar to the use of a written test as a hurdle as it allows police department to use other factors beyond written test scores to rank officers within the band.
19. Even the Defendant's own experts agree that within certain point spreads it is impossible to discern any difference in performance. Dr. Outtz testified that candidates should have been ranked in bands, because the test scores did not have sufficient reliability to use strict ranking. (15:59-62; 16:13-14). Dr. Jacobs, meanwhile, instructed the City to band the 2008 results. (Lopez Ex. 70). Thus, the City could have banded the scores allowing many more minorities to be considered for each round of hiring, and other components such as performance review could have been used to pick the better candidates. There is significant evidence that such a system would have less adverse impact.
20. Dr. Outtz testified that the reduction of adverse impact when banding can be improved by adding other selection criteria within the band by, for example, including a special interview process for the candidates within the band and by using prior performance in the department to break ties. (Lopez Tr. 15:19-20).
21. Dr. Fields also testified that adverse impact could be reduced through the use of banding. (Lopez Tr. 6:111).

C. The Valid Scientific Evidence Demonstrates That The Use Of Alternative Selection Procedures Has Less Adverse Impact On Minority Candidates.

22. All of the experts who testified in this case agree that there are a significant number of studies, social science papers, and validity reports which have carefully studied the question of whether certain selection techniques have more or less adverse impact on

minority candidates. These studies are based upon sound statistically significant scientific evidence.

23. There is no question that the consensus among industrial psychiatrists is that mean score difference between non-minorities and minorities is much higher on multiple-choice tests than on alternative testing components such as assessment centers, structures interviews, and oral boards. (12/16/14, p. 83).
24. Dr. Wiesen testified that alternatives such as structured interviews and assessment centers had from one-third up to one-half less adverse impact on minorities. (12/17/14, p. 84).
25. Dr. Fields testified that she had successfully reduced adverse impact in most of the examinations she developed which utilized multiple components. (Lopez Tr. 6:114)
26. Likewise, Dr. Outtz admitted that in the two places where he has devised police promotional exams using multiple components in the last 20 years – Detroit and Bridgeport – he has successfully reduced adverse impact. (Lopez Tr. 15:104-106).
27. Dr. Hough testified that she has devoted her whole career of 40 years to the “development and evaluation of personnel selection in promotion systems that have criterion- related validity with reduced or less adverse impact against minorities or protected groups.” 1/6/15, p. 16. She testified that she has developed promotional systems for, among others, Nynex, Microsoft, the Library of Congress, and the Federal Trade Commission and was successful in reducing adverse impact on minorities in each instance. (1/6/15, p. 17).
28. Dr. Wiesen testified about a chapter in a 2009 book called “Adverse Impact” which was written by Neal Schmitt and edited by Dr. Outtz which found that when

researchers gave a multiple-choice math test and then an open-ended math test, the D, or score difference, on Latinos was reduced by .66. ((Lopez Tr. 3:53-54).

29. Dr. Outtz testified in Lopez that the general expectation among industrial psychologists is that the use of an assessment center will reduce adverse impact on minorities. (Lopez Tr. 14:103).
30. Even when testing only for job knowledge, adverse impact can be reduced by using testing modes other than written, multiple-choice tests. (Lopez Tr. 3:51-55; 6:45-49).
31. Plaintiffs' expert Dr. Fields, who has developed more police promotional exams than the other experts in both Lopez and this case (Lopez Tr. 6:30, 34), testified that she has used well-developed multi-component testing to increase validity and decrease adverse impact in most jurisdictions where she has worked. (Lopez Tr. 6:114-15).
32. Dr. Fields testified that adverse impact could be reduced further through the use of weighting. Specifically, Dr. Fields testified that industrial psychologists had shown that overall adverse impact could be reliably reduced by giving less weight to components with the most adverse impact. (Lopez. Tr. 6:83)
33. According to a presentation given by Dr. Rick Jacobs to the International Personnel Management Association Assessment Council in 2000, Dr. Jacobs was able to reduce adverse impact in police examinations by adding testing components beyond a written cognitive ability test. 1/5/15, p. 40-43.
34. Dr. Outtz, in an article titled "The Role of Cognitive Ability Tests in Employment Selection" concluded: "Given that (a) cognitive ability tests can be combined with other predictors such that adverse impact is reduced while overall validity is increased; and (b) alternative predictors with less adverse impact can produce validity coefficients

comparable to those obtained with cognitive ability tests alone, sole reliance on cognitive ability tests when alternatives are available is unwarranted.” (Lopez Tr. 16:63-64). Dr. Outtz confirmed that opinion at the Lopez trial. (Lopez Tr. 16:64).

35. Adding components to an exam – i.e., expanding the range of critical KSA’s that are tested—tends to reduce adverse impact without any sacrifice to validity. (3:45-48; 16:67-68, 74-76; 17:129-30). More specifically, the evidence establishes that test procedures such as assessment centers, structured interviews, and video situational judgment tests tend to result in mean score differences that range from 0 to 0.55, while the mean score differences on multiple-choice job-knowledge tests range from 0.45 to over 1.0. (Lopez Tr. 3:57-60; 6:43-44; 16:80-81).
36. Dr. Outtz wrote an article concluding that there does not have to be a tradeoff between reducing adverse impact and increasing validity and that tests can achieve both goals at the same time by increasing the components and the dimensions tested for the job. (Lopez Tr. 16:67)
37. Dr. Outtz described what a meta-analysis is in Lopez, stating:

A meta-analysis is an accumulation of all the existing research data on a particular topic that you can find, that’s been published in any reputable source. It is the aggregation of that data such as you can draw conclusions as to the true state of a given phenomenon that you’re trying to study as opposed to relying on one or two individual studies from time to time in different journals.

In the field of industrial organizational psychology we perform meta-analysis periodically to see where we are based on the accumulated data up to that point and then it’s published. It also allows you to accurately make an estimate of the amount of error or variance associated with whatever phenomenon I am looking at. How far off am I? What’s the range in which I could possibly be off in making an estimate, for example, of the potential adverse impact of an assessment center? So it’s quite a valuable tool.

(Lopez Tr. 14:104).

38. Dr. Outtz testified that short-term memory tests have lower adverse impact. (Lopez Tr. 14:134)
39. In a 1997 article published in *Personnel Psychology*, the publication edited by Dr. Campion, by Dr. Paul Sackett and Dr. Jill Ellingson titled “The Effects of Forming Multi-Predictor Composites of Group Differences and Adverse Impact,” demonstrated that an exam with multiple components will reduce adverse impact if the components are properly correlated. (Ex. L; Lopez Tr. 15:7)
40. As Dr. Campion admitted, the meta-analysis performed by Schmidt and Hunter showed that adding components, such as assessment centers and work sample tests would increase the criterion-related validity of an exam that included a written test. (12/22/14, p. 106; Ex. 82)
41. Dr. Campion testified about a 1999 article by industrial psychologists Philip Bobko and Philip Roth published in *The Journal of Personnel Psychology*. (12/22/14, p. 129-30). The article included a meta-analysis that showed that the mean score difference between minorities and non-minorities on cognitive ability tests was 1.0 or, in other words, minorities scored, on average, one standard deviation worse than non-minorities. (12/22/14, p. 130).
42. The Bobko-Roth meta-analysis also found that the difference in job performance for minorities was only .45. (12/22/14, p. 133-34). Dr. Campion testified that this study showed, and all other studies have predicted, that the job performance of minorities is two-fold better than non-minorities based on their written test scores would indicate. (12/22/14, p. 134-35).

43. That is, as Dr. Campion agreed, “if you only use a multiple choice job knowledge test, you are going to see bigger differences in scores than you would expect to see if you actually looked at how such minorities were doing on the job at the Boston Police Department as lieutenants.” (12/22/14, p. 137).
44. In addition, there is evidence from Boston that non-multiple choice components have had significantly less adverse impact on minority candidates.
45. In 1985, Boston’s multi-component test had no adverse impact at the pass/fail level. (Ex. 71 (1991 Validation Report at p. 82); Lopez Tr. 8:100).
46. In addition, in 2002, the non-multiple choice assessment center components had significantly less adverse impact than the written exam, (Ex. 81) In fact the 2002 Lt.’s exam validity report authored by Dr. David Morris, concluded that there was no statistically significant adverse impact from the assessment center component. (12/22/14, p. 52; Ex. 81, p. 18, ¶ 5.1.)
47. Furthermore, Dr. Wiesen found that the D, or difference between minorities and non-minorities, on the 2002 written multiple choice exam was .55, while the D on the oral component was only .26. (Lopez Tr. 3:57). See also Exs. 80-81. In other words, the mean score difference between minorities and non-minorities was cut in half on the oral component. (Lopez Tr. 3:57-58).
48. In addition, even in 2014, the structured oral interview had significantly less adverse impact than the written test. (See Attachment to Callahan Affidavit).
49. Even without expert statistical analysis, it is easy to see that the oral component greatly reduced the adverse impact on minorities. The mean score for minorities on the 2014 written test (out of a possible high score of 36) was 25.58, while the mean score for

non-minorities was 28.15, for a mean score difference on the written test of 2.57 points. The mean score for minorities on the oral board (out of a possible 24) was 19.13, while the mean score for non-minorities was 20.19, for a mean score difference of 1.06. The mean score difference between minorities and non-minorities was, therefore, reduced by more than 50 percent on the oral board.

50. To demonstrate that the reduction in adverse impact is not simply due to the smaller maximum point total for the oral board, the reduction in adverse impact is clearly demonstrated when the scores for both components are converted to being out of 100 (by converting the scores into percentages of the total maximum score for each component). Out of 100, the mean score for minorities on the written test was 71.06, while the mean score for non-minorities on the written test was 78.19, for a difference of 7.13 points. Out of 100, the mean score for minorities on the oral board was 79.71, while the mean score for non-minorities was 84.13, for a mean score difference of 4.41. The mean score difference for the written test (7.13 points) was nearly two-times higher than the mean score difference on the oral board (4.41 points).
51. Further, a review of the briefs in Massachusetts Ass'n of Minority Law Enforcement Officers v. Abban, 434 Mass. 256, 262, 748 N.E.2d 455, 460 (2001), shows that in 1992, when the City of Boston used an exam that included a structured oral interview, there was a significant reduction in adverse impact. In fact, by the time that the City attempted to promote several minority candidates with scores of 84 over several white candidates with scores of 85, there had been no violation of the 4/5's rule. Abban, Brief of Boston Police Patrolmen's Association, 2000 WL 34607839, *7 (Mass.), at *7.

52. Many court cases have shown that the use of alternative testing procedures has significantly reduced adverse impact. For instance, Memphis achieved significant reduction in adverse impact in the 1990's by the use of its multi-component exam, and Chicago significantly reduced adverse impact by making 30% of its promotions based on a highly structured merit system. Johnson v. City of Memphis, 770 F.3d 464 at 475 (2014); Allen v. City of Chicago, 351 F.3d 306, 310 (7th Cir. 2003)
53. Dr. Rick Jacobs, in a presentation to the International Personnel Management Association Assessment Council in 2000 on his experience administering a promotional examination for Alabama state troopers, demonstrated that he was able to reduce adverse impact by adding multiple test components beyond a written test for cognitive ability. (1/5/15, p. 41).
54. In 2007, Dr. Outtz gave a presentation in South Africa on how to reduce adverse impact in governmental and private agencies based on data relating to various promotional tests given for fire lieutenants. (Lopez Tr. 16:77-78, 80). Dr. Outtz stated that African-American candidates suffered substantial adverse impact (a D of 1.3) on the written multiple-choice exams for fire lieutenant. (Lopez Tr. 16:81). Furthermore, Dr. Outtz concluded that the adverse impact on report writing was only one half of the impact on the multiple-choice test. Id. Dr. Outtz further found that the adverse impact on situational judgment tests was only one-quarter of that for written multiple-choice tests where there is one correct answer. Id. He also found that video situational judgment tests further reduced adverse impact by one half. Id.
55. In sum, there is overwhelming evidence that the use of alternative promotional testing mechanisms, which deemphasize M/C job knowledge will better assess the real skills

and abilities necessary to be a successful Lieutenant, while significantly reducing adverse impact.

D. Alternative Types Of Selection Procedures Having Less Adverse Impact.

56. There are a multitude of alternative selection procedures that have been shown to have less adverse impact while being predictive of good job performance as a police superior officer.

57. These include the following:

1. Performance review systems.

58. In 2002, the Boston Police Department decided to use a structured performance review system that was designed by their expert consultant firm at the time, Morris & McDaniel. (1/6/15, p. 88-90). The Department subsequently abandoned the idea solely because the unions challenged it before the Civil Service Commission. When he did abandon it, Commissioner Evans wrote an impassioned letter describing how essential a performance review system was and why it would have benefited the Boston Police Department. The letter stated:

It was my intention to provide candidates an opportunity to receive credit for the work they do day in and day out. . . . I believe that the best indicator of future performance is past performance, that what you do when you come into work counts. The best supervisors cannot always be identified solely by their performance on the written test and an hour in an assessment center. I wanted to give candidates credit for the work that they do on a daily basis. Quite simply, it is fairer to officers to factor in what they do on the job when they come to work.

(Lopez Ex. 194).

59. Commissioner Evans' lament was borne out by Sergeant Bruce Smith, who testified in this case about winning the Shattuck Award for his community involvement, his growing up on the streets of Boston, all of the excellent work that he had been doing,

his commendations, working as acting Lieutenant, but was never promoted to lieutenant. (12/17/14, p. 94-99).

60. Indeed, Commissioner Davis, and even current Commissioner Evans, gave testimony fully supportive of performance review systems. Commissioner Edward Davis testified that he sought to include a structured review system that would take account of past performance, which would measure abilities that were not measured on a written test, such as the ability to communicate. (1/5/15, p. 113). Commissioner Davis testified both in Lopez and in this case that being able to use past performance was extremely important in picking the best candidates, that many minority candidates would benefit from such a process because they were excellent minority candidates who simply didn't score well on a test. (1/5/15, p. 96-101, 108-09; Lopez Tr. 9:41-42).

Commissioners Davis and Evans testified how they used performance review in selecting the 20 or so non-Civil Service Deputy and Assistant Deputy positions that they are allowed to select outside of the Civil Service system. (1/5/15, p. 96; 1/7/15, p. 12-13). Davis testified that approximately 40% of his appointments were minorities using this system, that he has had excellent results, and that he has chosen people that would not have necessarily scored well on a test, but who have performed great at their job as Deputy or Assistant Deputy, a position over and above a Lieutenant. (1/5/15, p. 96)

61. Even current Commissioner Evans praised the fact that he had appointed a minority police Sergeant William Gross to be his second-in-command – Superintendent-In-Chief of the entire Boston Police Department and he had performed the job very well,

even though he had not even taken a lieutenant's promotional exam or had taken it and not done well enough to be promoted out of his sergeant's position. (1/7/15, p. 67-70).

62. Further, in a series of cases in Chicago, Illinois, the Chicago Police Department had agreed to a process where 30% of all of their promotions were based upon a structured performance review system and that process had resulted in "no" adverse impact. Allen v. City of Chicago, 351 F.3d 306, 310 (7th Cir. 2003).
63. Dr. Hough testified at length that certain systems like a performance review system that look at the individual's qualities, accomplishments, etc., are highly valid ways to make selections, and are known to have little or no adverse impact and are more valid than multiple choice tests. (1/6/15, p. 34-37).

2. Structured Oral Interviews

64. Another highly effective method of selecting promotional candidates is through the "structured oral interview" which is a well-recognized technique in the literature.
65. As was explained by many of the experts, under a structured oral interview, the questions are all arranged in advance and given to every candidate, the assessors and the experts debate what the appropriate answers would be, and the responses by the candidates, either by video or in person, are then evaluated based on a set group of standards. Further, as evidenced from the 2014 exam, structured oral interviews had significantly less adverse impact than did the written exam, and in 2002, the video scenarios/interviews were demonstrated to have no statistically significant adverse impact on lieutenant candidates. (Ex. 81, p. 18).
66. Significantly, Defendant's expert, Dr. Campion, is one of the leading proponents of the validity of the highly structured oral interview, and, in 1988, authored an article titled

“Structured Interviewing” which was the landmark study calling for the use of structured interviews and which described what steps could be taken to provide job-related structure to an interview. (12/15/14, p. 134; 12/16/14, p. 81). And since then, Dr. Campion has continued to published extensively on the subject with articles including “The Structured Employment Interview: Narrative and Quantitative Review of the Research Literature” published in Personnel Psychology in 2014 and “Are Highly Structured Job Interviews Resistant to Demographic Similarity Effects?” published in Personnel Psychology in 2010. (Ex. 73 at Ex. A).

67. Dr. Campion believes that a highly structured oral interview process has significant validity in selecting candidates, and he utilizes such a procedure in the attempt to reduce adverse impact. Indeed, in Dr. Campion’s 2010 article on structured interviews, he concluded that structured interviews could be conducted in a way that would reduce adverse impact that would be expected from racial similarities between interviewers and interviewees. (1/5/15, p. 40).
68. Dr. Wiesen, Dr. Outtz, and Dr. Fields all testified that these types of structured oral interviews are known to significantly reduce adverse impact when properly conducted. (Lopez Tr. 3:57-60; 6:43-44; 16:80-81).
69. The meta-analysis by Bobko and Roth found that structured interviews had a mean score difference of only .23. 12/22/14, p. 131.
70. Other studies have shown that, traditionally, the D on structured interviews is 0.1, much less than the D usually found on multiple-choice tests, which is closer to 1.0. Lopez Tr. 3:59- 60). Dr. Campion testified that an article by Neal Schmitt and Elaine Pulakos found that the D for tests for “verbal ability [such as cognitive ability test] was

1.03 while the D for structured interviews was 0.12 and the D for other measures, such as bio data and situational judgment tests had a D of 0.23. (1/5/15, p. 31, 33-34).

71. According to a 2009 article from the Human Resource Management Review by industrial psychologist Terese Macan, “Notably numerous meta analyses have revealed that structured interviews can display relatively high levels of validity without the adverse impact typically found in cognitive ability tests.” 1/5/15, p. 30. Dr. Champion agreed with this statement from the article. Id.

3. Situational Judgment Tests

72. Dr. Hough described situational judgment tests as “an assessment [in] which you describe a situation . . . and then you ask the applicant to describe what they would do, what they should do, a variety of response options.” (1/6/15, p. 23). Dr. Hough testified that situational judgment tests result in less adverse impact where if you ask the candidates to “describe what they should do” rather than what “would” they do. Id. Dr. Hough said that the reduction in adverse impact from situational judgment tests was clear from both her personal experience administering promotional examinations and also found to be the case by meta-analyses conducted by other industrial psychologists. (1/6/ 15, p. 24).
73. Dr. Hough further testified that the scientific research showed that situational judgment tests have even less adverse impact where the scenario is presented by videotape. 1/6/15, p. 24-26.
74. In Lopez, Dr. Outtz agreed that video-based tests resulted in significant reductions in adverse impact. (Lopez Tr. 16:64-65).

75. Dr. Wiesen, meanwhile, testified that the Pulakos-Schmitt article found that, where a situational judgment test was presented in a paper-and-pencil test and then was presented by video, the D, or discriminatory impact, was reduced by .74, which Dr. Wiesen concluded was a “huge reduction.” (Lopez Tr. 3:53-54).

76. Dr. Outtz described how the video test could be administered:

You can have the video and then have a person give an oral response to a panel as to how they would handle that particular situation, or you can have a video and have them choose between a number of strategies as to how they would handle that particular scene on the video, or you can have a video and have them speak into an audiotape saying what they would do, or you can have them watch the video and have a videotape of what they say they would do.

(Lopez Tr. 16:65-66).

77. In another scholarly article by several respected industrial psychologists, including Britt De Soete, Filip Lievens, Janneke Oostrom, and Lena Westerveld, published in the International Journal of Selection and Assessment in 2013, concluded: “At a practical level the present findings combined with the predictive validity evidence found in previous studies suggest that the constructive response multimedia test may be a valuable alternative predictor in [diverse] applicant settings.” And they further concluded that “This might be particularly relevant for police force selections as previous studies have revealed rather low predictive validity coefficients for cognitive ability tests in police contexts.” 1/5/15, p. 48.

III. PROPOSED CONCLUSIONS OF LAW ON LESS DISCRIMINATORY ALTERNATIVES

78. Even if somehow the court were to find the Lieutenants’ exam job-related and consistent with business necessity, even when used as a strict rank order device, the plaintiffs may still prevail by showing that they were equally valid and less

discriminatory alternatives available. Bradley v. City of Lynn, 443 F. Supp.2d at 156. The burden of proving equally valid less discriminatory alternatives is subject to the same standard of proof as in civil cases – plaintiffs must show by a preponderance of the evidence; meaning that it is more likely than not that there are equally valid selection processes that would have likely had lower levels of disparate impact. Id. at 175. Plaintiffs “have no obligation to provide the exact floorplan, or to have proposed the alternative before the exam was given.” Id. It is the plaintiffs' burden to prove that there was “another available method of evaluation which was equally valid and less discriminatory.” Bryant v. City of Chicago, 200 F.3d 1092, 1094 (7th Cir. 2000). The Plaintiffs are not required to have proposed the alternative, rather they need only show that “the alternative was available.” This has been interpreted as meaning that the Defendants knew or should have known of such alternatives. Johnson v. City of Memphis, 770 F.3d 464, 473 (6th Cir. 2014). It is undisputed that by 2008, all of the alternatives discussed herein were well known to social scientists, and had been carefully studied.

79. Preliminarily, the plaintiffs emphasize that when the City decided to administer the 2008 exam based upon a strict multiple choice job knowledge test, it knew with virtual certainty that the test would have severe disparate impact. Here and in Lopez, experts for both sides repeatedly made the point that a multiple-choice exam would result in the most adverse impact on minorities. nt. (12/15/14, pp.113-114; 12/18/14, pp.45-46, 70; Lopez Tr. 2:82-84; Lopez Tr. 3:43-47, 55, 60; Lopez Tr. 10:30; Lopez Tr. 16:59; Lopez Tr. 17:110). Dr. Outtz stated in writing that “the multiple-choice cognizability, or job-knowledge test, has the highest level of adverse impact.” (Lopez Tr. 3:44-46)

(emphasis added). Another expert emphasized, “[o]ne of the effects of using a written multiple-choice job-knowledge test is that you’re virtually assured of adverse impact.” (Lopez Tr. 3:47). Moreover, in Lopez, Boston actually conceded that the 2008 sergeant’s exam had adversely impacted minority candidates. The City knew, therefore, that, of all the various test designs that it could possibly utilize, a rank-order multiple-choice job knowledge test that accounted for virtually all of the exam score would have the highest and most severe adverse impact of any test and promotional process imaginable.

80. Thus, it is hard to understand how a court could conclude that there are no alternatives that would lessen the adverse impact of the promotional process, when the process used by the City was known in advance to have the highest level of adverse impact possible and rather low validity because it only tested for some job knowledge, not the critical skills and abilities, and attributes necessary to be a police lieutenant.
81. Because every expert that testified readily agreed that the addition of job-related components which test for additional skills and abilities always increases, not decreases, and because the multiple choice test has never been shown to have any significant correlation with job performance as a police superior officer, and does not test for most of the skills necessary for the job, Plaintiffs have clearly met their simple burden of demonstrating that the utilization of a multi-component promotional process would have been equally valid, if not more so.
82. Moreover, Plaintiffs have demonstrated that given the already documented proficiency of the candidate to obtain a high score on a job knowledge test when they became Sergeants, the fact of their performing on the job for many years (since that is a

prerequisite to sit for a lieutenants' exam) and the "range restriction" that results from this fact, the Plaintiffs have demonstrated by a preponderance of the evidence that utilizing a written multiple choice examination as a hurdle/qualifying device, and not as a scored rank order device, if combined with utilizing other components that test for skills and abilities not testable by a written exam, would increase not decrease validity. At the very least it would have equal validity while reducing adverse impact.

A. The City's Attempted Anecdotal Evidence That Utilizing Multi-Components Does Not Reduce Adverse Impact Is Not Based Upon Any Valid Evidentiary Showing.

83. Presumably, the City here will do attempt what it tried to do in Lopez and claim that because it had not obtained better results in 2002 and 2014 with respect to utilizing multi-components, the plaintiffs cannot demonstrate that alternative testing procedures would have less adverse impact. The problem with this argument is that it has been explicitly rejected by at least two circuit courts of appeal, based on evidentiary principles. In both Allen v. City of Chicago, 351 F.3d 306, 315-16 (7th Cir. 2003), and Johnson v. City of Memphis, 770 F.3d 464, 475 (2014), the courts held that the parties could not rely on the isolated results of a past examination process, in an effort to persuade the court that such evidence would predict future results for a test. Thus in both cases (Allen and Johnson), the plaintiffs attempted to demonstrate that their proposals would be "equally valid and less discriminatory" by relying on the past success those processes had had in prior tests in the Chicago and Memphis police departments respectively.
84. In Johnson, the officers proposed eliminating the written job skills test from the process so as to give full weight to merit review boards. The district court had accorded

“considerable significance” to the testing process used in 1996 that involved a job simulation which resulted in no adverse impact. The 1996 process was found to have “no adverse impact.” Id. at 475. But the Sixth Circuit held:

Yet as the Seventh Circuit explained in Allen – and we agree – past practice alone does not suffice. . .the past success of a specific testing process predicts but does not establish success in future applications.

Id. at 475.

85. Here, the City of Boston will claim that because the results of the 2002 exam were not as they had hoped, they would have no reason to believe that in 2008 there were less discriminatory alternatives. This argument suffers from the same defect that the Second and Seventh Circuit rejected since anecdotal evidence of one exam constructed in 2002 is not evidence. Indeed, as plaintiffs have shown, the non-written components of the exam did in fact have dramatically less adverse impact. Indeed, the 2002 assessment center had no statistically significant adverse impact on minorities. (Ex. 81, p. 18). The same is also true of the 2014 exam. In 2014, the structured oral interview had significantly less adverse impact than the written exam.
86. Accordingly, the City cannot purport to demonstrate that there were no less discriminatory alternatives by relying on one prior instance and one future instance. It is simply anecdotal evidence and not having any evidentiary value. Rather, the Court must value empirical social science studies of the effect of various test components and their relationship to adverse impact. On this question, virtually every study out there, and discussed in Lopez and here, demonstrates that utilizing non-written components will decrease adverse impact. Further, since using the written test as a qualifying device has been endorsed by many of the experts in this case, and has not been shown

to decrease validity, a final testing process that would have these other less discriminatory components, while still keeping the written test as a qualifying device, is to a reasonable likelihood (indeed is almost certain), going to have less adverse impact.

B. The Plaintiffs Have Demonstrated Equally Valid And Less Discriminatory Alternatives That Could Have Been Used.

87. There was testimony in this case that even when there are components of a promotional process that have lesser adverse impact, if they are combined with scores from a written multiple choice test that has significant adverse impact, the final result may not result in a substantial improvement of the adverse impact selection ratio. Essentially that is because the known adverse impact effect of the written M/C test will not likely be substantially ameliorated as long as that written score is being used as a substantial component, when scored on a rank order basis. Indeed, the 2002 exam demonstrates this point as there is only a slight improvement in the selection ratio for minorities because the written test was still accounted for over 40% of the score, and had severe adverse impact. Likewise, in 2014, there was severe adverse impact in the written exam, and although the structured oral interview had significantly less adverse impact, it is apparent that it will not result in a substantially reduced adverse impact because of the heavy weight that the City continues to place on the written test.
88. The Plaintiffs are not arguing for the elimination of the written M/C test as was the case in Allen v. City of Chicago. While that might substantially reduce adverse impact, assuming the selections are made from another component known to have lesser adverse impact, it does not permit a court to conclude that the final result was “equally valid.” Allen v. City of Chicago, 351 F.3d at 313.

89. Rather, Plaintiffs assert that the obvious solution is to combine a written multiple choice job knowledge test – used as a hurdle/ qualifying test with an appropriately set cut off score and carefully constructed by qualified experts in such a way as to reduce adverse impact, and to utilize and score other important components that test for the supervisory skills and abilities necessary to be a good police lieutenant, like communication, situational judgment, ability to take command, etc., all of which cannot be tested by a written exam, but all of which can be tested by structured oral interviews, assessment centers and the like.
90. If the written multiple choice test had been a hurdle, not a rank order device, many minorities would have been in the promotion pool. There would be a huge increase in minorities who could now be considered for promotion. But when the written M/C test is used as a rank order device, minorities generally sit at the bottom of the written test rank order list and never get real consideration. If that written hurdle then allowed passing candidates to take additional components such as the assessment center or structured oral interview, performance review, and other similar types of tests, the result would clearly be an equally or significantly more valid test (since more attributes were being examined), without any decrease in validity by reason of the written test being used on a qualifying basis. This is because, as stated above, there is no real evidence that police lieutenants who score higher on a written test perform better on the job than others who simply are determined to have successfully passed such an exam. Moreover, since these are lieutenants who have already successfully passed the sergeants' exam, they have already shown that competence. The non-written components would have significantly less adverse impact, as demonstrated by study

after study, and the inclusion of the additional components, and the different attributes they would test, ensures that there would be no less validity.

91. Based on the above, the plaintiffs have demonstrated that if in 2008 the City of Boston had used the written test as a hurdle, not a rank, and then combined it with well-designed components that are known to have less adverse impact, and test for the interpersonal and organizational attributes needed for the job, they would be equally valid and less discriminatory.

CONCLUSION

For these reasons, and the reasons stated in their earlier brief on adverse impact and validity, the Plaintiffs respectfully request that the Court issue findings of fact and rulings of law that the Defendant has engaged in disparate impact racial discrimination, in violation of Title VII and M.G.L. c. 151B.

Respectfully submitted,

BRUCE SMITH, et al.
By their attorneys,

/s/ Benjamin Weber
Harold Lichten, BBO#549689
Benjamin Weber, BBO#673736
Lichten & Liss-Riordan, P.C.
729 Boylston Street, Suite 2000
Boston, MA 02114
(617) 994-5800
hlichten@llrlaw.com
bweber@llrlaw.com

Stephen Churchill
Fair Work, P.C.
192 South Street, Suite 450
Boston, MA 02111

(617) 607-3262
steve@fairworklaw.com

DATED: May 4, 2015

CERTIFICATE OF SERVICE

I certify that a copy of the foregoing document has been served on counsel for all parties through the ECF system on this date.

DATED: May 4, 2015

/s/ Benjamin Weber
Benjamin Weber