

CS

**COMMONWEALTH OF MASSACHUSETTS
CIVIL SERVICE COMMISSION**

SUFFOLK, ss.

**BOSTON POLICE SUPERIOR
OFFICERS FEDERATION, et al,**

v.

Docket No. I-02-606, et al¹

**BOSTON POLICE DEPARTMENT
AND THE HUMAN RESOURCES DIVISION**

Appellants' Attorneys:

Harold L. Lichten, Esq.
Pyle, Rome, Lichten, Ehrenberg &
Liss-Riordan, P.C.
18 Tremont Street, Suite 500
Boston, MA 02108
(FOR: B.P.S.O.F.)

Joseph G. Donnellan, Esq.
Rogal & Donnellan, P.C.
43 Charles Street
Needham, MA 02494-2905
(FOR: Kenneth MacMaster & Trent
Holland)

Paul T. Hynes, Esq.
Angoff, Goldman, Manning,
Wanger & Hynes, P.C.
45 Bromfield Street
Boston, MA 02108-4106
(FOR: Arthur Stratford)

Respondent's Attorney

Mary Jo Harris, Esq.
Morgan, Brown & Joy LLP
200 State Street
Boston, MA 02109-2605
(FOR: City of Boston)

¹ Additional Case Nos. are B-02-476; B-02-477; B-02-478; B-02-479; B-02-480; B-02-481; B-02-482; B-02-483; B-02-484; B-02-485; B-02-486; B-02-487; and B-02-488. Also, I-02-846, I-03-113, I-02-86 and I-02-845 are cases that we are considering here because of the issues of fact common to them.

Michelle Heffernan, Esq.
Deputy General Counsel
Commonwealth of Massachusetts
Human Resources Division
One Ashburton Place
Boston, MA 02108

Commissioner:

John J. Guerin, Jr.

DECISION

Pursuant to the provisions of G.L. c. 31, §§ 2(a) and 2(b), the Appellants and the Boston Police Superior Officers Federation (hereafter the "Federation"), et al., challenge the Boston Police Department's (hereafter "Respondent" or "the Department") 2002 promotional examination for the ranks of lieutenant and captain, as the examinations were established by the Human Resources Division (hereafter "HRD"). The Appellants and the Federation seek an investigation by the Civil Service Commission (hereafter "Commission") to review and resolve questions as to whether the examinations and, specifically, the scoring of those examinations constituted a fair test of the abilities for the positions in question as required under G.L. c. 31, §§ 16 and 24.² Some hearings on the matters were held in 2003 by then-Commissioner Robert Tierney, Esq. Commissioner Tierney subsequently left the Commission and, after several attempts by the parties to effectuate an agreement resolving the matters proved unsuccessful, a series of hearings were held at the offices of the Commission on May 9 and May 10, 2006, and January 17 and June 19, 2007. The Commission's investigation consisted of hearing sworn testimony and accepting documentary evidence in order to make findings of fact and conclusions. Seven (7) audiotapes were made of the hearings in 2006 and 2007. All

² The Respondents filed a Motion to Dismiss on the grounds that the Commission lacked jurisdiction to hear the matter. The Appellants filed an Opposition to the Motion and a Surreply Brief. The Respondents filed a Reply to the Appellants' Opposition. The Commission denied the Motion on January 26, 2006.

parties submitted proposed conclusions and remedies to the Commission thereafter, as instructed.

FINDINGS OF FACT UPON INVESTIGATION:

Based upon the documents entered into evidence during the investigation (Joint Exhibits 1-20) and the testimony of Sergeant Bernard Greene, Lieutenant Detective Arthur Stratford, Dr. Kathleen Lundquist and Dr. Joel Weisen, I make the following findings of fact:

1. In or about January 2002, the Department and the HRD announced plans to administer promotional examinations for the positions of captain, lieutenant and sergeant in the Department. (Exhibit 4)
2. A January 14, 2002 memorandum from the Department informed candidates taking the examination that the written knowledge component of the test would *not* use Rule 200 as a source document for the examination. The Department noted in the memo that due to the terrorist attacks of September 11, 2001, Rule 200 was being updated. (Exhibit 4)
3. Rule 200 covers Critical Incident Management, containing the Department's rules and regulations for response to and management of critical incidents. Situations such as setting up a perimeter, hostage situations, being a first responder, calling in for back-up and calling other emergency response agencies are covered in this comprehensive rule. (Exhibit 1)
4. Announcement #8440 of the Promotion to Police Lieutenant Examination included a multiple choice written examination component, a performance-based

assessment component made of two subcomponents: the assessment center³ and the Performance Review System (“PRS”), and an education and experience review component. (Exhibit 6)

5. The points apportioned to each component of the 2002 examination were announced on January 11, 2002 when the PRS was introduced to the candidates for promotion. The presentation announced that for the lieutenant exam, the written component would be worth 30 points, the performance based assessment worth 50 points and experience and education worth 20 points. For the captain exam, the written test would be worth 25 points, the performance based assessment worth 55 points and experience and education worth 20 points. When the City agreed to withdraw the PRS from the 2002 examination, it left the points allocated to each component unchanged. (Exhibit 18)
6. The Federation, the Boston Police Patrolman’s Association and the Boston Police Detectives Union challenged the Respondents’ inclusion of a PRS as part of the examination process and to apportion 20 points out of the total 100 points towards the result of this component. The challenge was made before this Commission. (Boston Police Superior Officers Federation, et al v. Boston Police Department and Human Resources Division, Commission Docket No. G-02-143)
7. On June 24, 2002, a decision was issued by the Commission whereby the Appellants voluntarily withdrew their appeal on Docket No. G-02-143 in exchange for the PRS component being eliminated. The Respondents then

³ An Assessment Center consists of a standardized evaluation of behavior based upon multiple inputs. Several trained observers and techniques are used. Judgments about behaviors are made, in major part, from specifically developed simulations. (Exhibit 10)

- apportioned the 20 points from the PRS component to the assessment center portion of the examination. (Id.)
8. In the summer of 2002, the promotional examinations were administered by the Department under a delegation agreement from HRD. The examinations were designed by a consulting firm, Morris & McDaniel, hired by the Department to develop promotional examinations for Sergeants and Lieutenants. (Exhibit 19)
 9. 178 candidates for Lieutenant took that examination and participated in two assessment center exercises. 34 candidates for Captain took that examination and participated in three assessment center exercises. The examinations were a rank-ordered selection strategy with all candidates being ranked on an eligibility list in order of score, from highest to lowest. (Exhibit 6)
 10. The 2002 assessment center exercises were videotaped and assessors from outside the Department scored the exercises. These assessors were trained on the exercises and the scoring process and used a 1-7 scoring scale. Three panels of assessors scored the captain's examination and four panels scored the lieutenant's examination. Each panel consisted of three (3) members. (Exhibit 3 and testimony of Lundquist)
 11. Two of the three assessment exercises on the lieutenant's examination, in the video exercises portion, involved Rule 200 of the Departmental rules and regulations. Scenario 2 involved a hostage situation and Scenario 3 involved a first responder situation. (Exhibits 2 and 3 and testimony of Greene)
 12. Sergeant Detective Bernard Greene filed a June 26, 2002 appeal of the oral assessment portion of the examination, specifically questions 2 and 3 of the

situational exercise. Greene wrote that those two questions dealt with scenarios that fell under Rule 200 and were thus invalid as Rule 200 was not to be used as a source document for the examination. (Exhibit 7)

13. The September 26, 2002 response to Greene from the HRD stated, in part, that:

“While it is unavoidable that a person’s knowledge may facilitate the candidate’s demonstration of assessed dimensions, scenarios 2 and 3 were not designed to measure and did not measure knowledge of Rule 200.” (Exhibit 7)

14. At the Commission’s hearing, Greene testified that based on the January 2002 memorandum he did not prepare or study for Rule 200 and did not expect to confront any questions or parts of the examination based on Critical Incident Management. He stated that as he covered the busy Mattapan and Dorchester area, he learned different things than officers assigned to other districts and that based on an officer’s assignment, he may or may not know how to respond to critical incidents. (Testimony of Greene)

15. Greene was a very credible, comfortable and thoughtful witness who offered direct and thorough answers in his testimony. (Id.)

16. Appeals B-02-477 and B-02-488 were filed by sixteen (16) individual police sergeants claiming that the two video exercise questions that allegedly involved Rule 200 were not proper because they had been told not to study it. Their claims were adopted and pursued by the Federation. (Administrative Notice of appeal)

17. Dr. Joel Wiesen (“Dr. Wiesen”), a consultant in the area of industrial psychology with a doctorate in measuring candidates for work related performance, established his credentials as an expert witness and testified on behalf of the

Appellants. Dr. Wiesen possessed an expert grasp of the subject matter and offered relevant anecdotal testimony to explain concepts of reliability as well as very detailed and thoughtful answers. (Exhibits 8, 9 and 19)

18. After the examination scores were issued, the Department announced that there would not be a pass/fail score for either the written component of the examination or the assessment center. There is no professional standard within the industry that requires the use of a cut score. Historically, 70% is the score used by HRD as a cut score in its test administration as a pass/fail point. (Testimony of Wiesen)
19. Candidates were permitted to seek a review of their assessment center score on one or both of the scores given on the exercises and have their score rescored by a different panel. They were informed that their score could go up, down or remain the same. (Exhibit 17 and testimony of Wiesen)
20. Of the 178 applicants who took the lieutenant's examination, 70 requested rescoring of the video exercises. The score did not change for 57 of these 70 applicants. Eleven of the rescored tests rose, while two declined. Dr. Wiesen argued that although some candidates did not appeal, it did not mean the examination was a reliable one. (Testimony of Wiesen)
21. Thirteen of the 34 candidates for captain requested rescoring of the video exercises. Ten of those scores were changed; one had a 22% change, another a 17% change and a third had close to 14%. (Exhibit 15)
22. Dr. Wiesen testified that pass/fail scores are sufficient to determine the competence of an individual candidate but cut-off scores trim larger pools of candidates. (Testimony of Wiesen)

23. Dr. Wiesen testified that assessment center scores are not normally subject to an appeal process. Dr. Wiesen further testified that the scoring of the examination was flawed and that the large changes on the rescoring, differences of over ten or twenty points, called into question the reliability of the examination. He stated his belief that various panels should assess similarly to one another and the scores on rescoring should not be as widely varying as they were. (Exhibit 15 and testimony of Wiesen)
24. Dr. Wiesen stated that there were no empirical data to suggest either that the assessment scores were reliable enough or that the various panels interacted to improve reliability due to the wide changes in grading. He testified that the process used for re-scoring appeared to be more subjective than empirical based. (Exhibits 13 and 14 and testimony of Wiesen)
25. Dr. Kathleen Lundquist (“Lundquist”), an Industrial Psychologist with a Ph.D. in Psychometrics, testified for the Respondents that she works to develop testing and her area of expertise is defense of selection procedures. Dr. Lundquist testified that she did not make any determination regarding the reliability of the written portion of the test but stated that, based on a scientific examination of the reliability of the scores, the assessment center scores were sufficiently reliable to meet professional standards. Lundquist offered very credible and professional expert testimony. (Testimony of Lundquist)
26. Dr. Lundquist stated in her Expert Report responding to Dr. Wiesen’s opinions that, “Reliability is a necessary characteristic of a good test. Unless a test is reasonably reliable or consistent in its assessment of a candidate’s qualifications

from one administration to another or from one scoring to another, it cannot be a valid basis for making employment decisions such as the promotional decisions at issue....”(Exhibit 20)

27. Dr. Lundquist disputed Dr. Wiesen’s testimony, stating that the score changes indicated “reasonably” good reliability. Dr. Lundquist also testified that rotating panels is a “best practice” as it insures panels are mixed with both easy and more difficult raters. (Testimony of Lundquist)

28. Also in her Expert Report responding to Dr. Wiesen’s opinions, Dr. Lundquist stated that, “Based on a scientific examination of the reliability of the scores, it is my opinion that the assessment center scores are sufficiently reliable to meet professional standards.” She stated that the re-scoring of the assessment center exercises resulted in score changes within accepted standards of reliability. She further opined that the PRS points should be allocated to the assessment center as both are considered to be performance based exercises, involving the demonstration or application of the competencies being measured. (Exhibit 20)

29. I found that both expert witnesses were competent and their testimony was compelling. Both were instrumental in assisting the Commission to better consider a seemingly complicated fact pattern in a more manageable scope. Their testimony is very much appreciated by the Commission and their individual professionalism is gratefully noted here.

CONCLUSION:

G.L. c. 31, § 16 mandates that: “Examinations shall fairly test the knowledge, skills and abilities which can be practically and reliably measured and which are actually

required to perform the primary or dominant duties of the position for the which the examination is held.” The standard for fairness is noted in Ash v Police Commissioner, 11 Mass. App 650, 418 N.E. 2d 622 (1981).

The investigation of the fairness of the 2002 promotional examination is within the Commission’s jurisdiction as basic merit principles assure that employees are protected against arbitrary and capricious actions. Moreover, issues of the overall fairness of an examination and the credit or proportion of credit accorded to individual segments on an examination are well within basic merit principles. In making its conclusions, the Commission relies on the assistance offered by experts Dr. Wiesen and Dr. Lundquist in order to determine issues related to the design, scoring and re-scoring of the 2002 examination utilized by the Department.

In the present case, the Appellants challenge the fairness and reliability of the 2002 promotional examinations. With regard to the written examination, the Appellants challenge only the lack of a pass/fail cut off score. Additionally, the Appellants allege that the examination was not valid based on the Rule 200 questions, the significant disparities after re-scoring and the apportionment of the unused PRS points.

The Commission notes Dr. Lundquist’s statement in her report: “Reliability is a necessary characteristic of a good test. Unless a test is reasonably reliable or consistent in its assessment of a candidate’s qualifications from one administration to another or from one scoring to another, it cannot be a valid basis for making employment decisions

such as the promotional decisions at issue....”(Exhibit 20) With this observation as a guideline, the Commission addresses each of the Appellants’ points individually.

Cut Off Score

The Appellants argue that failure to have a pass/fail cut off score for the written examination results in that part of the examination not being a fair and reliable test of the attributes necessary for the job and violates the civil service statute. The Appellants maintain that the Department’s failure to establish a passing grade for the written portion of the examination potentially allowed some officers to be ranked based on the examination and promoted without possessing the basic knowledge necessary for promotion. The Appellants’ argument here is not persuasive. Although common and a standard method to delineate competency, there is no legal requirement or professional standard that requires a pass/fail cut off score. All components of the examination were constructed in order to test the relative knowledge, abilities and skills of the candidates in accordance with § 16. Therefore, the Commission concludes that, in this case, lack of a cut off score did not make the examination unfair.

Reapportionment of Points

The Appellants assert that the twenty (20) points that had been designated for the PRS portion of the examination and moved to the assessment center by the Department were improperly allocated and should have been split equally between the remaining

portions of the examination. However, given the similarity in method and content of measurement between the PRS and the assessment center exercises, about which Dr. Lundquist credibly testified in support of the PRS points being allocated to the assessment center scores, the Department established just cause therefore. Additionally, all candidates had knowledge, as of January 2002, as to how points would be apportioned for the three measured components of these examinations. The Appellants' success in eliminating the PRS did not, and should not, have disturbed the overall scoring structure of the examinations.

Assessment Center Scoring

With regard to the reliability of the assessment center scores, the Commission must determine whether the Department proved that they were a reasonably accurate and reliable measure of ability, knowledge and skills relevant and material to the positions for which the assessment is being made. Dr. Wiesen opined that there were such large score changes upon the re-scoring of the examination that it was not reliable while Dr. Lundquist reached the contrary conclusion. While both these experts were knowledgeable and credible, we find that the assessment center scoring was sufficiently accurate and reliable. Thus, the Commission will not disturb the scoring or re-scoring of the assessment center exercises.

Rule 200 Issue

We now discuss the concerns relating to the Rule 200 issue. The Appellants requesting this investigation maintain that the Department breached basic principles of

fairness when it specifically instructed candidates not to study Rule 200 and then included two questions on the video exercises that related to Rule 200. The Department countered that, although those taking the test were told that Rule 200 was not on the reading list, that did not relieve them of the need to know how to respond if presented with scenarios that would assess their ability to respond to a critical incident during the assessment center exercises. However, as Greene testified, it is plausible that some officers would have had experience with critical incidents due to prior assignments while others would not. This discrepancy results in an unfair advantage for some candidates and is based not on their preparation for the examination but on their prior assignments.

Faced with this situation, the Commission cannot help but draw on experience with regard to preparing for significant and determinative examinations. In plain language: Being informed that something will not be on the test and then having it be tested is not fair.

Accordingly, the Commission finds that Rule 200 played a role in the scoring of the two of three video scenarios after officers were told not to study this Critical Incident Management rule for the examination. Although, as the Respondent argues, a promotional examination for police lieutenant or captain would reasonably be assumed to assess judgment in the face of an emergency when determining the fitness of a candidate for a supervisory position, the inclusion of the two video scenarios relating to Rule 200 breaches basic principles of fairness. Therefore, the Appellants can rightly claim that

their rights under c. 31 were prejudiced in this regard, through no fault of their own.

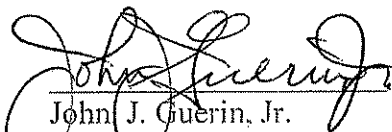
Chapter 310 of St. 1993 of the Acts and Resolves of Massachusetts provides that,

"If the rights of any person acquired under the provisions of chapter thirty-one of the General Laws or under any rule made thereunder have been prejudiced through no fault of his own, the civil service commission may take such action as will restore or protect such rights, notwithstanding the failure of any person to comply with any requirement of said chapter thirty-one or any such rule as a condition precedent to the restoration or protection of such rights."

For the above reasons, the Appellants' appeals on Docket Nos. I-02-606, et al are hereby *allowed with respect to the Rule 200 issue only* and hereby *dismissed* with respect to other issues of the appeal.

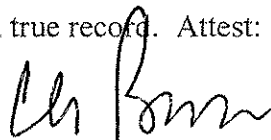
Pursuant to the powers inherent in Chapter 534 of the Acts of 1976, as amended by Chapter 310 of the Acts and Resolves of 1993, the Commission hereby grants equitable relief to those Appellants *who actually filed examination appeals to the Commission pursuant to G.L. c. 31, § 2(b) whose appeals are so noted by Docket Numbers beginning with the letter "B" herein*, by instructing that any Appellant who so requests may have his or her examination rescored with the exclusion of the disputed video scenarios and a new score calculated. If, based upon this rescoring, an Appellant would have received a score that equals or is greater than the score of anyone who was promoted based upon that 2002 examination, then such Appellant, if not yet promoted, shall be granted Chapter 310 relief, so called, and have his or her name placed at the top of the current promotional list until such time as his or her name has been reached for consideration for a vacant position.

Civil Service Commission


John J. Guerin, Jr.
Commissioner

By vote of the Civil Service Commission (Bowman, Chairmen; Guerin, Marquis, Taylor, and Henderson) on January 31, 2008.

A true record. Attest:



Commissioner

A motion for reconsideration may be filed by either Party within ten days of receipt of a Commission order or decision. A motion for reconsideration shall be deemed a motion for rehearing in accordance with M.G.L. c 30A, section 14(1) for the purpose of tolling the time for appeal.

Any party aggrieved by a final decision or order of the Commission may initiate proceedings for judicial review under section 14 of chapter 30A in the superior court within thirty (30) days after receipt of such order or decision. Commencement of such proceeding shall not, unless specifically ordered by the court, operate as a stay of the Commission's order or decision.

Notice to:

Harold L. Lichten, Esq.

Joseph G. Donnellan, Esq.

Paul T. Hynes, Esq.

Mary Jo Harris, Esq.