

**COMMONWEALTH OF MASSACHUSETTS  
CIVIL SERVICE COMMISSION**

**SUFFOLK, ss.**

One Ashburton Place: Room 503  
Boston, MA 02108  
(617) 727-2293

**WAYNE G. CLARKE,**  
*Appellant*

**Docket No: B2-16-44**

v.

**HUMAN RESOURCES DIVISION and  
BOSTON POLICE DEPARTMENT,**  
*Respondents*

Appearance for Appellant:

Wayne G. Clarke, Pro Se

Appearance for Human Resources Division:

Michael Downey, Esq.  
Human Resources Division – Room 211  
One Ashburton Place  
Boston, MA 02108

Appearance for Boston Police Department:

Jaclyn Zawada, Esq.  
Boston Police Department  
One Schroeder Plaza  
Boston, MA 02120

Commissioner:

Paul M. Stein

**DECISION**

The Appellant, Wayne G. Clarke, appeals to the Civil Service Commission (Commission),<sup>1</sup> pursuant to G.L.c.31,§24, to appeal a review by the Massachusetts Human Resources Division (HRD) of the markings of his answers to essay questions in the In-Basket portion of the September 2014 Assessment Center Examination for Police Lieutenant with the Boston Police Department (BPD), which review HRD conducted pursuant to G.L.c.31,§22 and the Commission's prior Decision (Exh.5) in Clarke v. Human Resources Division, 29 MCSR 1 (2016) (Clarke I). A hearing of the appeal was held at the Commission's Boston Offices on December 2, 2016 and was digitally recorded with copies of the CD provided to the parties.<sup>2</sup>

---

<sup>1</sup> The Standard Adjudicatory Rules of Practice and Procedure, 801 CMR §§1.00, *et seq.*, apply to adjudications before the Commission with Chapter 31 or any Commission rules taking precedence.

<sup>2</sup> If there is a judicial appeal of this decision, the plaintiff in the judicial appeal becomes obligated to use the CDs to supply the court with the stenographic or other written transcript of the hearing to the extent that he/she wishes to challenge the decision as unsupported by the substantial evidence, arbitrary and capricious, or an abuse of discretion.

Eleven exhibits were received in evidence (Exhs. 1 through 8 and Confidential<sup>3</sup> Exhs. C9, C10/PH-C10A and C11). HRD submitted a Proposed Decision on March 24, 2017 and the Appellant made post-hearing submissions on March 2, 2017 and March 31, 2017.

## **FINDINGS OF FACT**

Based on the Exhibits entered into evidence and the testimony of the following witnesses:

*Called by HRD*

- Briana Ward, HRD's Director of Test Development,

*Called by the Appellant:*

- Wayne G. Clarke, Appellant

and taking administrative notice of Clarke I and relevant matters filed in this appeal, as well as pertinent law and reasonable inferences from the credible evidence, I find that the preponderance of evidence establishes the following facts:

1. The Appellant, Wayne G. Clarke, is a permanent BPD Sergeant with a civil service seniority date of November 23, 1989. (*Administrative Notice [Clarke I]*)
2. In April 2013, HRD entered into a Delegation Agreement through which BPD was delegated authority to design and administer a departmental promotional examination for the positions of BPD Sergeant, Lieutenant and Captain. (*Exh. 4; Testimony of Director Ward*)<sup>4</sup>

---

<sup>3</sup> The disclosure of information in discovery and at the hearing was subject to Commission rulings intended to address HRD's concerns that certain documentation, although relevant to its review of the Appellant's In-Basket examination scoring, contained information that was proprietary to the outside test consultant who designed and administered the In-Basket examination, as well as certain other information concerning the substance of the examination and the scoring of the examination exercises that could compromise future examinations were they to be publically disseminated. Accordingly, the Appellant was required to execute a form of Protective Order approved by the Commission that limited his access to these materials and required his review of other documents only under a controlled environment. The Appellant declined to execute the Protective Order and objected to the Commission's ruling regarding the limitations on his access to information. Thus, save for two modifications, the hearing proceeded without certain evidence that would have been available if the Appellant elected to execute the Protective Order. (*Exhs.6,8 & 9, Confidential Exh.C10; Administrative Notice [Discovery Orders];Hearing Colloquy*)

<sup>4</sup> Director Ward also testified in another In-Basket Test appeal brought by a BPD Lieutenant who challenged his markings on the 2014 BPD Captain's Examination, which appeal was heard by the Commission on December 1, 2016 and for which the Commission's Decision is also issued today. Wilbanks v. HRD, CSC No. B2-16-47 (Wilbanks II) The Commission takes administrative notice of the testimony provided by Director Ward in both this appeal and in Wilbanks II)

3. Pursuant to the Delegation Agreement, BPD retained, with HRD's approval, the firm of EB Jacobs as its consultant, which designed and administered a BPD Police Lieutenant's examination that consisted of three components, one of which was an In-Basket Test (administered on September 6, 2014) that counted 20% toward the candidate's total examination score.<sup>5</sup> (*Exhs. C10 & C11; Administrative Notice [Clarke I]; Testimony of Director Ward*)

4. The In-Basket Test was a one-day, "open-book" "Assessment Center" style examination in which the candidate was asked to assume the role of a newly promoted Lieutenant and to provide written, essay-style responses to job-related problems (exercises) typical of those a Lieutenant might encounter. Prior to the examination, candidates received a Candidate Preparation Guide describing the examination components, suggestions for preparation strategies and sample examination materials. At the test center, candidates received Oral Instructions and a Background Information Packet that included such documents as calendars, personnel rosters and organizational charts, as well as a series of memos, reports and other correspondence typical of those documents that might come across a Lieutenant's desk. Candidates had approximately three hours to review the background materials and prepare a written Response Booklet to demonstrate how they would handle the problems presented in the information packet. (*Exh. C11; Administrative Notice [Clarke I]; Testimony of Director Ward*)

5. The Response Booklet was evaluated by two-member panels (superior officers in police departments outside the Commonwealth) trained by EB Jacobs. Each assessor independently scored the responses in four categories, also referred to as "competencies", for each of the two scenarios (exercises) presented by the test, using a nine point "Likert" scale (9 is high and 1 is

---

<sup>5</sup> The two other components of the examination were a Written Technical Knowledge Test (administered on June 28, 2014), comprising 36% of a candidate's total score and an Oral Board Test (administered over the course of two days, October 8 and 9, 2014), comprising 24% of the total score. Sgt. Clarke's scores on these two other components are not now challenged in this appeal. (*Exhs. 7 & C10; Administrative Notice [Clarke I]*)

low), as well as assigned “overall” combined scores for each competency, i.e., Written Communication, Interpersonal Interactions, Analyzing and Deciding, Managing Activities. (*Exh. C10; Administrative Notice [Clarke I]; Testimony of Director Ward*)

6. The assessors were trained to assign scores for each competency by first choosing one of three broad categories to describe the response and then fine-tune the choice by assigning one of the three numerical scores within that category to best describe how many (i.e., all, most or a majority) of the pre-defined assessment criteria for that competency (contained in the assessors’ training manual and information provided to candidates) that the candidate’s response satisfied. The categories were: Highly Effective [7 to 9], Moderately Effective [4 to 6] and Ineffective [1 to 3]. (*Testimony of Ward; Administrative Notice [Wilbanks II]*)

7. EB Jacobs examined the results of the In-Basket Test by panel and determined that there were “significant differences” in the average competency raw scores by panel. After initial “standardization” of the raw scores by panel, EB Jacobs arrived at revised average raw scores for each of the four In-Basket competencies (Written, Communication, Interpersonal Interactions, Analyzing and Deciding and Managing Activities) that were computed to five decimal places, with the lowest possible “standardized” score being 1.07083 and the highest possible score being 8.13335 for each specific competency. These “standardized” competency scores were then totaled to arrive at the Overall In-Basket Component Score for each candidate that was further “rescaled/weighted” and totaled to arrive at a Final EBJ Examination Score for each candidate’s In-Basket Test. (*Exh. C10; Administrative Notice [Clarke I]; Testimony of Director Ward*)<sup>6</sup>

---

<sup>6</sup> “Standardization” to account for differences in the panels’ raw “competency” scores is distinguished from the Rescaling/Weighting process later used to adjust the Overall Component scores “to carry the intended test weight” for each test component (Written Exam, In-Basket, Oral Board) necessary to arrive at the Final EBJ Examination Score. According to EB Jacobs’ Feedback Report, unlike standardization, Rescaled/Weighed adjustment to overall component scores did not affect the candidates rank on that component and that conclusion was not disputed. (See *Exh. C10, pp. 2 & 4; Finding of Fact No.9, infra*)

8. EB Jacobs’s “standardization” of the panel’s average raw scores of for each competency in the In-Basket Test employed data and a proprietary algorithm that EB Jacobs declined to disclose to HRD or to the Commission. (Exh. C10; Testimony of Director Ward)

9. EB Jacobs’s “Rescaled/Weighted” scores for the Overall In-Basket Component (and the other two test components) were derived by using a formula that was provided to candidates in the “Score Notice and Feedback Report” provided to each candidate. This report stated:

“It is important to note that standardization and rescaling scores for any examination **component** does not affect your rank on that examination **component**. For example, if you received the third-highest **Overall Technical Knowledge Score**, you would also receive the third highest **Rescaled/Weighted Technical Knowledge Score**. Standardizing and rescaling is an accepted scoring practice and is used for examinations such as the college SATs.”

(Exh. C10) (emphasis in original; *emphasis added*)

10. According to EB Jacobs, after performing “standardization” and “rescaling/weighting”, the total overall In-Basket Test scores ranged from 13.49082 (the lowest score received by any candidate) to 20.0000 (the highest possible score), with a Mean Score (average) of 16.92648. Candidates at the 25% percentile received a total overall In-Basket Test score of 15.47910. Candidates at the 75% percentile received a total overall In-Basket score of 18.59857.(Exh. C10)

11. Sgt. Clarke duly registered for and took the 2014 promotional examination for BPD Police Lieutenant. He received the following competency scores from the two assessors who graded each of the two exercises in the In-Basket Test:

<u>Assessor 1</u>	<u>Written</u>	<u>Interpersonal</u>	<u>Analyzing/Deciding</u>	<u>Managing</u>
Exercise A	8	4	4	4
Exercise B	8	4	4	4
Overall	8	4	4	4
<u>Assessor 2</u>	<u>Written</u>	<u>Interpersonal</u>	<u>Analyzing/Deciding</u>	<u>Managing</u>
Exercise A	7	4	4	4
Exercise B	7	3	3	3
Overall	7	3	3	3

(Exh. C9; Testimony of Director Ward)

12. After “standardization” and “rescaling/weighting”, Sgt. Clarke received a Final Overall Examination Score of 79.78072, composed of the following examination component scores:

<b><u>EXAMINATION COMPONENT</u></b>	<b><u>SCORE</u></b>
Technical Knowledge	29.10014
In-Basket Test	14.61752
Oral Board Test	18.83306
FINAL EBJ EXAMINATION SCORE	62.45072
EDUCATION AND EXPERIENCE POINTS	17.33
VETERAN’S POINTS OR 25-YEAR SENIORITY POINTS	0 _____
<b><u>FINAL OVERALL EXAMINATION SCORE</u></b>	<b><u>79.78072</u></b>

(*Exh. C10*)

13. Sgt. Clarke’s final overall examination score, after rounding to a whole number as provided by HRD Personnel Administration Rules, caused him to be ranked on the current BPD Police Lieutenant’s Eligible List with a passing score of 80. (*Exh. C10; Administrative Notice [PAR.07(4)]*)

14. On January 9, 2015, Sgt. Clarke duly sought HRD’s review and recalculation of the scores he received on the In-Basket, Education and Experience (E&E) and Twenty-Five Year Experience components of his BPD Lieutenant’s examination. (*Exh. 1*)

15. After his request was denied, Sgt. Clarke appealed to the Commission which allowed his appeal in part, ordering that HRD conduct a further review of his In-Basket Test score and add additional points to his E&E score to which he was entitled by virtue of having earned a college degree. In its Decision in *Clarke I*, the Commission upheld HRD’s determination that Sgt. Clarke had not established that he had twenty-five years of service and ruled that Sgt. Clarke’s claim for additional points on that component was correctly denied. (*Exhs. 2, 3 & 5*)

16. As a result of the Commission’s Decision in *Clarke I*, HRD recalculated Sgt. Clarke’s E&E score, producing an adjusted Final Overall Examination Score of 80.98072 which, after rounding, equates to a whole number score of 81. (*PH Exh. C10A*)

17. Also, in compliance with the Commission's Decision in *Clarke I*, HRD's Director of Test Development, Briana Ward, was assigned to perform a review of the markings of Sgt. Clarke's In-Basket Test. (*Exh. 6; Testimony of Director Ward*)

18. Director Ward holds a Master of Arts degree in Industrial & Organizational Psychology, which she received in 2011. She began her employment at HRD in 2014. Her responsibilities include overseeing the creation, review and updating of civil service examinations administered by HRD. Her experience at HRD primarily has involved the administration of statewide written entry-level and promotional civil service examinations, such as examinations for appointment as an Environmental Police Officer (EPO), EPO Sergeant and Lieutenant, and examinations for appointment to the Massachusetts Department of Correction, Correction Officers I, II (Sergeant) and III (Lieutenant). Prior to the review that is the subject of this appeal, Director Ward had no direct experience in constructing, marking or reviewing the markings of essay question examinations, either at HRD or elsewhere. (*Testimony of Director Ward*)

19. Neither HRD nor Director Ward were involved in the administration of the 2014 BPD Lieutenant's examination, save for responsibility to calculate and review points awarded to candidates on the E&E component of the examination. Director Ward has general familiarity with the duties and responsibilities of a municipal police officer in Massachusetts, but has no direct training or experience in police work. Her review of the markings of Sgt. Clarke's In-Basket Test scores relied on this general familiarity, in consultation with EB Jacobs, the BPD's Legal Unit and other HRD professionals, along with the documentation that EB Jacobs provided to her, as well as her formal education and experience in Industrial & Organizational psychology. (*Testimony of Director Ward*)

20. Director Ward's review of the markings of Sgt. Clarke's In-Basket Test began with the receipt of all the documents from EB Jacobs that Director Ward deemed necessary for her review.<sup>7</sup> These documents can be categorized into four broad categories:

- A. In-Basket Test Assessor Training Manual – Director Ward used this document to gain insight into how EB Jacobs trained the assessors to evaluate the candidate's In Basket responses according to the nine-point rating scale (a form of Likert rating scale commonly used in personnel evaluations).
- B. In-Basket Oral Instructions to Candidates (read to candidates at the time of the test) and the Candidate Preparation Guide for the In-Basket (written guide provided to candidates in advance) – These documents, particularly the written guide, provided Director Ward with many relevant details about the examination process.
- C. In-Basket Test Background Information Booklet/Blank In-Basket Test Candidate Response Booklet – These documents contain the test exercises that each candidate received at the test center, the information associated with two scenarios to be reviewed and absorbed prior to making his/her written responses in the response booklet, and a set of four questions that the candidates were expected to specifically address as to each scenario.
- D. Sgt. Clarke's In-Basket Response Booklet, Test Consolidation Form (assessor's raw score sheet) and Notice & Feedback Report – Copy of Sgt. Clarke's written responses to the two In-Basket exercises, the assessors' scores (1 to 9) for each competency and overall score for each exercise, and a feedback report provided to Sgt. Clarke containing

---

<sup>7</sup> As noted earlier, most of these documents were not provided to Mr. Clarke and were not introduced in evidence due to Sgt. Clarke's decision not to execute the required Stipulated Protective Order. The description of the documents is based on the testimony of Director Ward as well as limited confidential documents that HRD did provide during the hearing. (*Exhs. 6, 8, C9 & C10; Testimony of Director Ward*)

his “standardized” and “rescaled/weighted” scores and final exam score, along with a narrative of assessor feedback on areas of strength and areas that needed improvement.

- E. Six Candidates’ Completed In-Basket Test Response Booklets, Test Consolidation Forms (assessor’s raw score sheet) and Notice & Feedback Reports – The completed response booklets, assessors’ scores, and feedback reports for six sample candidates selected by EB Jacobs, two of whom predominately were scored Highly Effective (7 to 9), two Moderately Effective (4 to 6) and two with scores in the Ineffective range [1 to 3]

*(Exhs. 6, 8, C9 through C11; Testimony of Director Ward; Administrative Notice [Wilbanks II])*

21. Director Ward familiarized herself with the relevant background information provided by EB Jacobs, including the Assessors Training Manual, the Oral Instructions to Candidates, the Candidate Preparation Guide, the Test Background Information Booklet and the Blank Response Booklet. She then proceeded to read the Response Booklets of each of the six sample candidates (without knowing their assessors’ scores) and, applying her own judgment, mentally scored each of the four competencies for each candidate’s responses on the two exercises, using the nine-point Likert rating scale and the rating methodology provided in the Assessors Training Manual described above. (She did not write down the scores but kept them “in her head”.) She then compared the scores she mentally assigned to the six candidates with the raw scores assigned by the assessors to those same candidates. She testified that, in every case, her assigned scores “aligned” with the assessors’ scores, meaning that, generally, they fell within the same broad category and within a point of the numerical raw scores assigned by the assessors. *(Exh. C13; Testimony of Director Ward; Administrative Notice [Wilbanks II])*

22. Director Ward performed the same review of Sgt. Clarke’s Response Booklet. She scored his responses and then compared her numerical scores to the scores assigned by the

assessors. She testified that, in every case, her scores “aligned” with the assessors. (*Exh. C9; Testimony of Director Ward*)

23. Director Ward also reviewed the Notice and Feedback Reports for the six candidates and for Sgt. Clarke, but did not analyze the data they contained. She did not analyze the “standardization” methodology that EB Jacobs employed to reconcile the “significant differences” in the assessors’ scores by panel. Specifically, she was not provided with the identity of the assessors, how many different assessors and panels were used or the data or the algorithm that EB Jacobs employed in the standardization process, so she did not know the details of how that standardization was performed. Therefore, she could not replicate or perform such analysis herself or specifically explain how the candidates’ final scores shown in the Notice and Feedback Reports were computed. (*Exh. C10; Testimony of Director Ward*)

24. By letter dated February 24, 2016, Director Ward notified Sgt. Clarke that she had completed her review of the markings of his In-Basket Test. The letter “came to the conclusion that your test data aligned with the rating scales and were appropriate given the scores of the sample candidates that were reviewed. According to the information provided by EB Jacobs I concluded that you were scored accurately on the In-Basket Test for the Lieutenant examination.” (*Exh. 6*)

25. This appeal duly ensued. (*Exh. 7*)

#### Applicable Civil Service Law

The process for HRD review and appeal to the Commission to challenge the results of a civil service examination are currently contained in G.L.c.31, Sections 22 through 24 and follow a distinctly different statutory path from other forms of civil service appeals from HRD actions (or inactions). See, e.g., G.L.c.31, §2(b) (Commission is granted power and duty “[t]o hear and

decide appeals by a person aggrieved by any decision, action, or failure to act by the administrator, except as limited by the provisions of section twenty-four relating to the grading of examinations)” (*emphasis added*) Here, Sgt. Clarke’s appeal to the Commission challenges HRD’s review of the marks on his “In-Basket” Test scores, as the Commission directed in *Clarke I.*<sup>8</sup> Thus, this appeal invokes that part of G.L.c.31, §22 through §24 which provides:

“. . . [A]n applicant may request the administrator [HRD] to conduct . . . a review of the marking of the applicant’s answers to essay and multiple choice questions . . . .” G.L.c.31, §22, ¶2 (*emphasis added*)

“Within six weeks after receipt of a request [for a §22 review], the administrator [HRD] . . . shall conduct such review, render a decision, and send a copy of such decision to the applicant. If [HRD] finds that an error was made in the marking of the applicant’s answer to an essay question . . . [HRD] shall make any necessary adjustment to correct such error.” G.L.c.31, §23 (*emphasis added*)

“An applicant may appeal to the commission from a decision of [HRD] . . . relative to (a) the marking of the applicant’s answers to essay questions . . . . no later than seventeen days after the mailing of the decision of [HRD]. . . . [T]he commission shall conduct a hearing . . . , render a decision, and send a copy of such decision to the applicant and [HRD]. . . .” G.L.c.31, §24 (*emphasis added*)

#### Analysis

This appeal comes to the Commission with little relevant prior judicial or Commission precedent. Challenges to the marking of essay questions have been extremely rare. The only judicial precedent to address the subject is more than fifty years old and construed a statute that differed significantly from the current version of G.L.c.31, §22 through §24. *See Clarke I.*<sup>9</sup> As

---

<sup>8</sup> Similar issues are presented in *Wilbanks II.*

<sup>9</sup> Sgt. Clarke’s present appeal is to be distinguished from a “fair test” appeal, separately authorized by G.L.c.31, §22 through §24, which is not the type of appeal presented here. A “fair test” appeal challenges the examination, in whole or in part, on the grounds that it did not constitute “a fair test of the applicant’s fitness actually to perform the primary or dominant duties of the position for which the examination is held . . . .” G.L.c.31, §22, ¶4; G.L.c.31, §24(b). A fair test appeal may involve, for example, claims that questions were erroneously framed, covered subjects as to which applicants did not have notice, or other irregularities in the test procedure that gave undue advantages or disadvantages to some applicants over others. *See, e.g., DiRado v. Civil Service Comm’n*, 352 Mass. 130 (1967) (applicants not given equal opportunity to use drawing aids); *Boston Police Super. Officers Federation v. Civil Service Comm’n*, 35 Mass.App.Ct. 688 (1993) (video performance component, an essential part of the examination, was tainted by test administrator’s conflict of interest) *See also O’Neill v. Civil Service Comm’n*, MICV09-0391 (2009), *aff’d*, 78 Mass.App.Ct. 1127 (2011) (Rule 1:28) (time to bring “fair test” appeal); *Swan v. Human Resources Div.*, CSC No. B2-15-182 (2015)(same)

this record showed, until this appeal, HRD’s Director of Test Development had never performed a review of the marking of an essay question examination.

In Clarke I and Wilbanks I, the Commission ruled that BPD’s In-Basket Test, but not the Oral Board Test, was an “essay” test subject to HRD “review” and appeal to the Commission under G.L.c.31,§22 through §24.<sup>10</sup> The Commission also made an initial determination that rejected HRD’s contention that the scope of the required HRD review should be limited to a mathematical “computational” exercise, but, rather, the Commission determined that it required a more substantive, thorough review of the markings of the essay questions to determine whether the scores represented a fair assessment of the candidate’s performance as compared to others, or were demonstrably “arbitrary or wholly devoid of reason.” The Commission left it to HRD to establish specific review procedures that would meet the required substantive review standard. The Commission left open for future consideration what parameters applied to the Commission’s jurisdiction and “hearing” in a further appeal under Section 24 of G.L.c.31 of HRD’s decision rendered after making such a “review”.

The parties do not now dispute the Commission’s determination that HRD had the obligation to conduct an in-house substantive review of the markings of Sgt. Clarke’s In-Basket Test. The parties differ, however, on the standard of review that must be applied upon appeal of HRD’s review to the Commission. The parties also differ as to whether or not the process employed by HRD to conduct that review meets the required statutory standard.

#### The Commission’s Standard of Review

As noted in the Commission’s Decisions in Clarke I and Wilbanks I, the standard of review to be applied upon appeal from HRD’s review of the markings of an essay question under the

---

<sup>10</sup> The Commission’s decision in Wilbanks I, which determined that the Oral Board Test was not an essay test, is pending judicial appeal.

current version of Chapter 31, Sections 22 through 24, has not been definitively determined. Case law decided under prior versions of the civil service law holds that “the marking of an examination answer is a finding of fact” as to which the Director of Civil Service (a predecessor to HRD) is given “broad discretionary powers” to exercise “judgment as to . . . proper grading of . . . examinations” and that, on appeal the Commission is “vested with similar” fact-finding powers, which findings of fact should not be disturbed unless “clearly shown to be arbitrary or devoid of logic and reason.” See Ferguson v. Civil Service Comm’n, 344 Mass. 484, 487-88 (1962)(upheld Commission’s overturning Civil Service Director’s decision to decline to increase the marking of an applicant’s answer to an essay question “through mistake”, finding that applicant “in his answer . . . stated the crucial fact that a sentence to state prison could not be imposed” and “should receive full credit for his answer”); Barry v. Civil Service Comm’rs, 323 Mass 431 (1948) (upheld Commission’s decision to increase marks on certain answers by one applicant, finding that “the applicant submitted authority to substantiate the correctness of the answers”, and declined to increase the marks of another applicant, rejecting his contention that different standards were used in the marking of the examination papers of the two applicants)

Civil service law then provided, as to examination appeals, that “no decision of the director [of civil service] relating to an examination mark shall be reversed and no such mark changed unless the commission finds that it was through error, fraud, mistake or in bad faith, and in each case of reversal of such decision or change in marking the specific reasons therefor shall be stated . . . . St. 1945, c. 725, §1 (*emphasis added*). In 1971, the Attorney General noted this prior version of the examination review and appeal requirement in a 1971 opinion, citing Moore v. Civil Service Comm’n, 333 Mass. 430, 434 (1956):

“One of the subjects with which the special commissions and the Legislature were especially concerned was that relating to examinations. It is apparent . . . that the making

up and grading of examinations were to be primarily administrative functions to be performed by the director and that the appellate jurisdiction of the commission related to examination marks was to be more restricted than it was in other matters.”

Op.Atty.Gen., Nov. 19, 1971, citing 333 Mass. at 434 (*emphasis added*).

Shortly thereafter, the legislature completely rewrote the examination review statutes, restricting the review by the Director of Civil Service to requests that the “computation of [an applicant’s] general average mark be checked for error” and completely eliminating all right of appeal to the Commission. St.1973, c.320, §§1 through 5.

Then in 1974, the legislature enacted a major administrative restructuring of the civil service system. The division of civil service was severed from the Commission, abolished and replaced by the division of personnel administration (DPA), headed by a Personnel Administrator (the “administrator”) reporting to and appointed by the secretary of Administration and Finance. All technical, executive and administrative functions of the division of civil service were transferred to DPA and the administrator was substituted for the position of director of civil service in all respects set forth in Chapter 31. The Commission became an independent quasi-judicial agency that retained its investigatory and appellate authority, including rule-making approval, over the actions of the administrator (and otherwise) as provided by Chapter 31, but no longer exercised indirect supervision and control over the functions transferred to DPA. St. 1974, c.835.

A year later, the legislature restored the authority it had removed in 1973, enacting statutory language closer to the present version of Sections 22 through 24 that, among other things, restored authority to the Personnel Administrator (the successor to the Director of Civil Service) to review the marking of essay questions and restored the Commission’s authority to hear examination appeals from such “findings of the administrator relative to grading of answers to essay questions”. St. 1975, c. 358, §§3, 4.

In Lincoln v. Personnel Administrator, 432 Mass. 208 (2000), the Supreme Judicial Court addressed the question whether, under the restructured (current) civil service statutes, the plaintiffs (who challenged alleged changes in the way the scores on their 1996 firefighter examination had been determined) could appeal for a “hearing” directly to the Commission from the marking of his/her examination, or whether they were first required to seek a “review” of their scores by the Personnel Administrator of DPA (now HRD). The plaintiffs had argued, and the Superior Court had agreed, that, since the Personnel Administrator designed, administered and scored the examination in the first instance, to provide an “additional review after the examination” was futile, as it “would make him the judge of his own challenged unfairness, something that . . . the Legislature would not have intended without much more explicit language.’ ” Id., 432 Mass. at 210. The SJC disagreed, and upheld the Commission’s dismissal of the petitions for failure to exhaust administrative remedies by first seeking the Personnel Administrator’s review, specifically rejecting the argument that such review was a futility.

“Nor is this an instance in which exhaustion is excused as futile. . . . It is true that the statute does require the personnel administrator to review his own action in response to a petition from an applicant. However, as the personnel administrator designs, administers, and scores the examinations, he possesses expertise in regard to the grading and weighing of the examination. As the statute is designated, the initial review by the personnel administrator allows him to apply that expertise, determining whether there has been a mistake, or an issue has been overlooked, that can be easily corrected before an eligibility list is certified. Therefore, the personnel administrator is the most familiar with the examination and is best able to respond to applicants who have raised questions regarding the grading of the examination. . . .”

Id., 432 Mass. at 212-13.

Finally, the Commission must be mindful of the reality that, in distinct contrast to the role that the Personnel Administrator (and HRD) historically played (noted in Lincoln) as the authority who “designs, administers and scores” all civil service examinations), more recently, as in the case of the 2014 BPD Lieutenant’s examination, HRD plays a peripheral role in many

examinations, delegating the authority to design, administer and score the examinations to a private consultant selected by the municipality or state agency who requests the examination. Moreover, civil service examinations are now generally limited to public safety positions and are no longer given for most civil service titles. Thus, while HRD remains, in theory, the technical expert in matters of civil service examinations, the institutional scope of that expertise, in fact, is not what it once was.

I have carefully considered the relevant legislative history described above, as well as the history of the examination process as it has evolved over time. I conclude that G.L.c.31,§24 does not mandate that the Commission put on “assessors’ hats’ and conduct a “de novo” hearing to supersede HRD’s “review” of the grading of an essay question. Rather, it suffices that the Commission’s role in an appeal from the review of marking of an examination question hew to its traditional quasi-judicial appellate oversight of HRD’s “actions or inactions, so as to ensure that HRD’s decision is based on the type of “impartial and reasonably thorough” review that has been required of the Commission’s review of other HRD “actions or inactions under G.L.c.31, Section 2(b), and that HRD’s conclusions are neither “arbitrary” nor “devoid of logic”, but, rather, supported by a preponderance of credible evidence.

This standard affords sufficient discretion to HRD to perform its duty to conduct the required first-level review, subject to appropriate evidentiary scrutiny by the Commission consistent with the Commission’s well-defined core responsibilities, as developed in related case law, to serve as an appellate check on HRD’s compliance with all civil service law and rules, generally. See, G.L.c.31, §2(b) & §5(a). See generally, Police Dep’t of Boston v. Kavaleski, 463 Mass. 680, 688-89 (2012) citing Massachusetts Ass’n of Minority Law Enforcement Officers v. Abban , 434 Mass. 256, 259 (2001); Brackett v. Civil Service Comm'n, 447 Mass. 233, 241 (2006) and cases

cited; Beverly v. Civil Service Comm'n, 78 Mass.App.Ct. 182, 187 (2010); Leominster v. Stratton, 58 Mass.App.Ct. 726, 727-28 (2003). See also Mayor of Revere v. Civil Service Comm'n, 31 Mass.App.Ct. 315, 321 (1991); Selectmen of Wakefield v. Judge of First Dist. Ct., 262 Mass. 477, 482 (1928) With this standard in mind, I turn to the evidence introduced to support HRD's conclusion that Mr. Clarke's In-Basket Test was correctly marked and to justify HRD's decision to decline to adjust it upward.

#### *The Sufficiency of HRD's Review*

First, Mr. Clarke argues, as a procedural matter, that HRD prevented a meaningful examination by the Commission of HRD's review of his In-Basket Test by unlawfully withholding from him and the Commission substantially all of the materials Director Ward used to perform the HRD review of his In-Basket Test scores. He contends that, therefore, the Commission must find that HRD acted arbitrarily and without reason or logic. He asserts that, as a result of HRD's withholding of relevant evidence, by default, he is entitled to be awarded full credit – i.e., 20.0000 points, the “Highest Score” for the In-Basket Test component.

Mr. Clarke's procedural argument deserves serious consideration. On the one hand, he makes the point that any materials that HRD used to review the administration and scoring of his In-Basket Test are relevant to his appeal to the Commission brought to challenge that review. On the other hand, HRD presents legitimate reasons why these materials are not “public records” and are protected from general disclosure. In particular, as Mr. Clarke acknowledges, the Massachusetts Public Records law excludes from the definition of a “public records”: (1) “questions and answers, scoring keys and sheets and other materials used to develop, administer or score a test, examination or assessment instrument . . . . intended to be used for another test, examination or assessment instrument” and (2) “trade secrets or commercial . . . information

voluntarily provided to an agency for use in developing government policy and upon a promise of confidentiality, but this shall not apply to information submitted as required by law or as a condition of receiving a governmental contract or other benefit”. G.L.c.4,§7(26)(g) & (l).

The applicability of the public record exclusions to the materials used by HRD relevant to this appeal is not entirely clear. HRD claims that the examinations designed by EB Jacobs for BPD’s 2014 promotional examinations “may” be used for future examinations and the disclosure of the materials could become a source of generating unfair advantage in future examinations. In addition, HRD represents that EB Jacobs claims that its test design, administration and scoring process contains proprietary information and that public disclosure of the information would compromise its competitive position in the testing business. Although these assertions were not definitively established, and may well be doubted, after consideration of the position of the parties in this appeal, I approved a proposed Stipulated Protective Order that accommodated the interests of Mr. Clarke to access the information he needed to prepare his appeal as well as the interests of HRD, BPD and EB Jacobs to limit the dissemination of that information. The proposed Stipulated Protective Order struck an appropriate balance between these disputed, competing interests. Ultimately, Mr. Clarke elected not to sign the proposed Stipulated Protective Order and he must bear the consequences of that decision. Thus, non-disclosure of information that would otherwise have been provided had Mr. Clarke executed the proposed Protective Order, alone, does not justify awarding him full credit on his In-Basket Test.

Second, Mr. Clarke also questions Director Ward’s qualifications to conduct a review of his In-Basket Test. He demonstrated that she was not familiar with federal litigation upholding

challenges to other public safety examinations designed and administered by EB Jacobs.<sup>11</sup> He also contends that Director Ward misunderstood the “subjective” nature of an essay examination under Massachusetts Civil Service Law. I am not persuaded that these discrepancies, taken together with Director Ward’s academic credentials, as well as the evidence as a whole, discredit her testimony or warrant rejection of her review as a matter of qualification or bias, despite her limited actual experience with HRD.

Third, Mr. Clarke challenges the selection of the comparative sample of six other In-Basket Test candidates’ results provided by EB Jacobs that purportedly represented two candidates who achieved an overall score in EB Jacob’s “Highly Effective” range [7 to 9], two who scored “Moderately Effective” [4 to 6] and two who scored “Ineffective” [1 to 3]. He contends that the sample was too small and that EB Jacobs would have been able to skew the sample by picking examples as they saw fit. The evidence, however, does not come close to establishing either of these technical objections, and they remain no more than mere speculation.<sup>12</sup>

Fourth, I considered whether HRD’s review was “impartial and reasonably thorough” and whether its conclusions were supported by a preponderance of credible evidence or, rather, were “arbitrary” or “devoid of logic”. I conclude that, save in two respects, HRD has met this test.

The materials that Director Ward examined prior to conducting her review of Mr. Clarke’s In-Basket test included all the information that reasonably appeared necessary and appropriate to HRD’s review under the circumstances of this case.<sup>13</sup> This preliminary review included the

---

<sup>11</sup> Mr. Clarke cites two civil rights cases, one of which resulted in judgment for plaintiff firefighters in Akron, OH, and one case involving police officers in Pittsburg, PA that was settled. Howe v. City of Akron, 801 F.3d 718 (6<sup>th</sup> Cir.2015), on remand, 2016 WL916701 (N.D.Ohio 2016); Foster v. City of Pittsburg, 2015 WL 11112431 (W.D.Pa. 2015)

<sup>12</sup> The six-candidate sample is less than 5% of the total applicants who took the BPD 2014 Lt. Exam. Expert testimony would be helpful if this issue sample size is raised in the future.

<sup>13</sup> Mr. Clarke did not direct HRD to any specific “books or publications” in support of his request for a review of his examination responses as provided by G.L.c.31, §22, ¶5.

material Director Ward needed to obtain a general understanding of the subject matter covered by the In-Basket Test, as well as the components that the assessors were asked to evaluate (Written Communications, Interpersonal Interactions, Analyzing and Deciding and Managing Activities), as well as the detailed instructions provided to the assessors for scoring the responses of each candidate. Director Ward then proceeded to review the six sample test booklets along with the scoring reports of the assessors for each sample, which I find was also a reasonable approach to gaining further understanding about how the assessors applied the assessment tools they were provided to specific test responses. After completing this educational process, Director Ward then applied the information she had obtained to her specific, independent review of Mr. Clarke's In Basket Test responses and scores. In the final stage of the process, Director Ward concluded that the raw scores awarded by the assessors to Mr. Clarke's In-Basket Test responses were "aligned" within the same range as indicated by the raw scores she assigned to his responses in her independent review. Up to this point, although less than perfect, HRD's review was both "independent" and "thorough"<sup>14</sup>.

I do conclude, however, that HRD's review falls short in two related respects: i.e., (1) HRD overlooked a critical factor essential to a "reasonably through" review which directly affected Mr. Clarke's final scores, namely, the "standardization" that substantially reduced his raw scores, purportedly to account for "significant differences" in the scoring by different panels; and (2) by limiting the review to a comparison of the raw scores, Director Ward's conclusion that her own scoring "aligned" with the assessors' scores becomes problematic.

---

<sup>14</sup> I note that most, but not all of the elements that the assessors were required to score, related to written communications, interpersonal, and management skills that are not unique to supervisory duties of a police officer. Some criteria did involve analysis of law enforcement specific subject matter as to which the assessors (all law enforcement personnel) would be intimately familiar, but Director Ward would not. It would certainly have added an additional level of confidence in the review had Director Ward had that knowledge or made further inquiry, but, given her otherwise thorough review, I do not find that deficiency, alone, grounds to discredit her conclusions.

The evidence is not disputed that HRD’s review did not make any effort to evaluate the effect of EB Jacobs’s “standardization” of the raw “competency” scores into what actually became Mr. Clarke’s In-Basket Test Average Ability Scores for each competency. As a result of the “standardization”, all of Mr. Clarke’s scores were reduced significantly. His Written Communications score was reduced from the “Highly Effective” range (7s & 8s) to 4.89741, dropping him as barely “Moderately Effective”. His Interpersonal Interactions, Analyzing & Deciding and Managing Activities competency average scores dropped from average of 3.5 to 2.77785, 2.70838 and 2.77905. After these adjustments, Mr. Clarke’s Overall In-Basket Component Score totaled 13.16258. Based on the reduction in his scores, Mr. Clarke’s Overall In-Basket Test score placed him below the 25% percentile of all candidates’ final scores, and within one point of the lowest score of all “Ineffective” candidates.

In the absence of “standardization”, the average of the two assessors’ overall competency scores given to Mr. Clarke would have totaled 18.0000, placing him in line with the assessors’ “Highly Effective” marks for Written Communication and borderline “Moderately Effective/Ineffective” in all other categories, based on the following raw scores:

<u>Overall Score</u>	<u>Written</u>	<u>Interpersonal</u>	<u>Analyzing/Deciding</u>	<u>Managing</u>
Assessor 1	8	4	4	4
Assessor 2	7	3	3	3
Average	7.5	3.5	3.5	3.5

In other words, all of Mr. Clarke’s adjusted individual and average competency scores differ from the raw scores given by the assessors (and, presumably, the scores given by Director Ward which she testified “aligned” with them.) Both assessors, and presumably Director Ward, rated him “Highly Effective” in Written Communications, but his adjusted score dropped him 2.6 points to “Moderately Effective”; Assessor 1 rated him borderline “Moderately Effective” in all other competencies on both exercises and Assessor 2 rated him borderline “Moderately

Effective” on all other competencies in one of the two exercises; yet again, his “standardized” scores dropped him to the low end of “Ineffective” on all of those other competencies.

Director Ward took no steps to ascertain how EB Jacobs adjusted these assessors’ raw scores. EB Jacobs did not provide her with the formula that was used to compute these adjustments nor did she receive any information that explained what “significant differences” were noted by EB Jacobs in the assessors’ scoring that purportedly required “standardization”. I also note that Director Ward’s review was premised on her conclusion that her “competency” scores “aligned” with the raw scores assigned by Mr. Clarke’s two assessors, and with the assessors’ raw scores assigned to the six other test subjects. Thus, the conclusion that Director Ward found Mr. Clarke’s raw scores were appropriately aligned with hers simply cannot be squared with the fact that EB Jacobs rejected these raw scores and significantly reduced them on some unexplained basis. At a minimum, HRD should have made inquiry into this discrepancy. By failing to do so, HRD did not meet its obligations to conduct a proper review of Mr. Clarke’s marks. HRD’s presumption that EB Jacobs apparently treats its algorithms and analysis used to justify “standardization” as proprietary does not alter this result. Therefore, the conclusion that Mr. Clarke’s In-Basket scores are marked correctly fails to pass the test of being reasonably thorough and supported by credible evidence but, rather, is arbitrary and devoid of logic or reason.

The final question that remains is what relief is appropriate. The Commission could remand this matter to HRD for another review consistent with this Decision. Alternatively, the Commission is authorized to order an adjustment warranted by the evidence presented to the Commission. Based on that evidence, an In-Basket Test score of 16.19017 would reflect the score Mr. Clarke would have received save for EB Jacobs’ unexplained “standardization” of raw

assessors' scores.<sup>15</sup> With this adjustment to Mr. Clarke's original In-Basket component score, plus all recalculated E&E points (18.53), his Final Overall Examination Score would be increased to 82.65337. After rounding, this score becomes the whole number 83 for purposes of placement on the eligible list. Such an increase in his rounded score would move Sgt. Clarke from 78<sup>th</sup> place to 63<sup>rd</sup> place on the current BPD Lieutenant Eligible List.

Prior to awarding Sgt. Clarke this relief, however, HRD should be afforded one further opportunity to establish, by a preponderance of evidence, a rational explanation for the standardization process applied by EB Jacobs and to explain to the Commission's satisfaction that: (1) without standardization of the raw competency scores on the Lieutenant In-Basket Test, Sgt. Clarke would have ranked exactly the same as he was with standardization (i.e. tied with eight others at 78<sup>th</sup> place) on the BPD Lieutenants Eligible list and (2) if the standardization did make a difference, the facts, including all data, upon which EB Jacobs relied to conclude that there were "significant differences" across panels that required standardization (explaining, for example, why the differences were attributed to panel scoring discrepancies, as opposed to the "luck of the draw", e.g., some panels happened to get more highly effective candidates and some got more ineffective candidates), the methodology that was used to make the standardization adjustments, and HRD's assessment that the facts and methodology support the need for the adjustment and the specific adjustments made.

---

<sup>15</sup> This calculation averages the two overall assessors scores for each of the four competencies [7.5, 3.5, 3.5,3.5] and inserts the total into the formula used by EB Jacobs to calculate the Overall Component Score [V] for the In-Basket Test:  $V = ((([18.0] - 20.129450/5.17357) * 1.78890) + 16.92648) = 16.19017$ . See *Exh. C10, p. 4*

## **CONCLUSION**

For the reasons stated above, the appeal of the Appellant, Wayne G. Clarke, is *allowed, in part*. It is hereby ORDERED that HRD shall:

A. Within 45 days of this decision, provide to the Commission (and a copy to the Appellant) a further submission on the issues of standardization of the raw competency scores on the 2014 BPD Lieutenant's In-Basket Examination, including (1) all facts that EB Jacobs relied (i.e. raw scores) to conclude that there were "significant differences" in scores by panel, (2) the methodology used to make the standardization adjustments to the raw competency scores; (3) a listing (anonymous) for each candidate whose raw scores were adjusted by standardization indicating each competency score prior to and after adjustment, (2) a listing (anonymous) for each candidate whose final ranking on the 2014 BPD Lieutenant's Examination would have been different, and if so what the difference would have been, had the raw scores on the candidate's In-Basket Examination not been standardized; (4) competent expert testimony from HRD or other person to establish, with reasonable professional certainty, that the standardization was employed and applied in a manner consistent with sound statistical and test administration standards.

B. Upon receipt of HRD's submission, the Commission may take such further action on this appeal as will appear necessary and proper, including, without limitation, reopening the hearing to conduct such further inquiry as required.

C. As an alternative to, and/or in the absence of the receipt of the response prescribed by Paragraph B, HRD shall, with 45 days of this Decision, take such action as necessary to increase the Appellant's EBJ In-Basket Examination Score to 16.19017, bringing his Final Overall Examination Score to 82.65337, and his rounded whole number score to 83, and to adjust his

place on the current BPD Lieutenant's eligible list accordingly, consistent with this Decision and civil service law and rules.

Civil Service Commission

/s/ Paul M. Stein

Paul M. Stein

Commissioner

By vote of the Civil Service Commission (Bowman, Chairman; Camuso, Ittleman, Stein & Tivnan, Commissioners) on July 20, 2017.

Notice to:

Wayne G. Clarke (Appellant)

Michael Downey, Esq. (for HRD)

Nicole I. Taub, Esq. (for BPD)

Either party may file a motion for reconsideration within ten days of the receipt of this Commission order or decision. Under the pertinent provisions of the Code of Mass. Regulations, 801 CMR 1.01(7)(l), the motion must identify a clerical or mechanical error in this order or decision or a significant factor the Agency or the Presiding Officer may have overlooked in deciding the case. A motion for reconsideration does not toll the statutorily prescribed thirty-day time limit for seeking judicial review of this Commission order or decision.

Under the provisions of G.L.c.31,§44, any party aggrieved by this Commission order or decision may initiate proceedings for judicial review under G.L.c.30A,§14 in the superior court within thirty (30) days after receipt of this order or decision. Commencement of such proceeding shall not, unless specifically ordered by the court, operate as a stay of this Commission order or decision. After initiating proceedings for judicial review in Superior Court, the plaintiff, or his / her attorney, is required to serve a copy of the summons and complaint upon the Boston office of the Attorney General of the Commonwealth, with a copy to the Civil Service Commission, in the time and in the manner prescribed by Mass. R. Civ. P. 4(d).