

**COMMONWEALTH OF MASSACHUSETTS
CIVIL SERVICE COMMISSION**

SUFFOLK, ss.

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

SEAN M. WILBANKS,
Appellant

Docket No: B2-16-47

v.

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

Appearance for Appellant:

James F. Lamond, Esq.
McDonald Lamond Canzoneri
352 Turnpike Road, Suite 310
Southborough, MA 01772-1756

Appearance for Human Resources Division:

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

Appearance for Boston Police Department:

Nicole I. Taub, Esq.
Jaclyn Zawada, Esq.
Office of the Legal Advisor
Boston Police Department
One Schroeder Plaza
Boston, MA 02120

Commissioner:

Paul M. Stein

DECISION

The Appellant, Sean M. Wilbanks, appeals to the Civil Service Commission (Commission), pursuant to G.L.c.31,§24,¹ 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 Captain with the Boston Police Department (BPD), which review HRD conducted pursuant to G.L.c.31,§22 and the Commission's prior Decision in Wilbanks v. Human Resources Division, 29 MCSR 22 (2016)

¹ 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.

(Wilbanks I). A hearing of the appeal was held at the Commission's Boston Offices on December 1, 2016 and was digitally recorded.² Eighteen exhibits were introduced in evidence (Exhs. 1 through 6 & 18; Confidential³ Exhs C7 through C11, C12A-C12F, C13A-C13F, C14A-C14F, C15 through C17). The Appellant and HRD submitted Proposed Decisions on March 24, 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,

and taking administrative notice of *Wilbanks I* and relevant matter 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, Sean M. Wilbanks, is a permanent BPD Lieutenant. (*Exh. 2*)
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. 18; Testimony of Director Ward*)⁴
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 Captain's examination that consisted of three components, one of which was an In-Basket Test

² Copies of the CDs of the full hearing were provided to the parties. 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.

³ Confidential Exhibits were introduced subject to a Discovery Order and a Stipulated Protective Order entered on August 30, 2016. (*Exhs. 5 & 6*)

⁴ Director Ward also testified in another In-Basket Test appeal brought by a BPD Sergeant who challenged his markings on the 2014 BPD Lieutenant's Examination, which appeal heard by the Commission on December 2, 2016 and for which the Commission's Decision is also issued today. *Clarke v. HRD*, CSC No. B2-16-47 (*Clarke II*)

(administered on September 6, 2014) that counted 19.2% toward the candidate's total examination score.⁵ (*Exhs. 2, C9, C14, C15, 18*]; *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 Captain and to provide written, essay-style responses to job-related problems (exercises) typical of those a Captain 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 roster and organizational charts, as well as a series of memos, reports and other correspondence typical of those documents that might come across a Captain'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. (*Exhs. 2, C8 through C11*; *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 three scenarios (exercises) presented by the test, using a nine point "Likert" scale (9 is high and 1 is low), as well as assigned "overall" combined scores for each competency, i.e., Written Communication, Interpersonal Interactions, Analyzing and Deciding, Managing Activities. (*Exhs. C7, C13, C16*; *Testimony of Director Ward*)

⁵ The two other components of the examination were a Written Technical Knowledge Test (administered on June 28, 2014), comprising 32% of a candidate's total score and an Oral Board Test (administered over the course of three days, September 24 through September 26, 2015) comprising 28.8% of the total score. Lt. Wilbanks' scores on these two other components are not now challenged in this appeal. (*Exhs. 2, C9, C15*)

6. The assessors' were trained to score each competency by first choosing one of three broad categories to describe the response – i.e., Highly Effective [7 to 9], Moderately Effective [4 to 6] and Ineffective [1 to 3] – and then to fine-tune the selection by assigning one of the three numerical scores within that chosen category that best described how many pre-defined assessment criteria for the chosen competency category that the candidate's response had satisfied (i.e., all, most or a majority). EB Jacobs' training manual for assessors specified nine specific test criteria (A –I) for the Written Communications component and ten criteria (A – J) for the other components (Interpersonal, Analyzing & Deciding and Managing). These criteria are also disclosed in the Candidate Preparation Guide. (*Exhs. C7 &C9; Testimony of Ward*)⁶

7. EB Jacobs' training materials also flagged “Potential Rating Errors” assessors were prone to commit and identified “Strategies for Minimizing Rating Errors”, such as:

- Letting global impression or one characteristic influence all judgments.
- Assuming candidate with characteristic “A” must have characteristic “B”.
- Tendency to be a “HARD” or “EASY” rater.
- Giving average or middle of the scale ratings to all candidates.
- Letting quality of other candidate(s) influence judgments about present candidate.
- Overly influenced by early or very late responses.
- Judge candidates “similar to ourselves” more favorably.
- Weigh unfavorable information too heavily

Assessors were counseled to minimize errors by:

- Reviewing all instructions before rating candidates
- Do not rely on memory - take notes
- Consider each candidate independently
- Consider each competency independently
- Complete review of each issue before making ratings
- Check list of “potential rating errors”

(*Exh. C7*)

⁶ According to an excerpt from the EB Jacobs In-Basket Training Manual introduced in evidence, assessors received “extensive training” during which they were “familiarized with the job duties and responsibilities of BPD captains”, were “trained to observe, record and evaluate candidates’ behaviors with respect to the abilities to be assessed”, and practiced evaluating candidates in “mock” assessment exercises. The Commission did not receive the entire Training Manual and heard no specific evidence as to what specific training actually occurred; neither the Commission, the parties nor Director Ward knew the identity or credentials of the assessors who scored Lt. Wilbanks or any other candidates. (*Exh. C7; Testimony of Ward*)

8. EB Jacobs averaged the two assessors' competency scores to arrive at an In-Basket Test Average Ability Score for each competency and, then, totaled the four average raw scores to arrive at an Overall In-Basket Raw Score. (*Exhs. C14 & C15; Testimony of Director Ward*)

9. EB Jacobs then "Rescaled/Weighted" the raw scores for the Overall In-Basket Component (and the other two test components) by a formula that was set forth 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. C15*) (emphasis in original; **emphasis added**)

10. According to EB Jacobs, after "rescaling/weighting", overall In-Basket Test scores ranged from 10.00309 (the lowest score any candidate received) to 19.2000 (the highest possible score), with a Mean Score (average) of 15.25095. Candidates at the 25% percentile received an overall In-Basket Test score of 13.82068. Candidates at the 75% percentile received an overall In-Basket Test score of 16.51034. (*Exh. C15*)

11. Lt. Wilbanks took the 2014 promotional examination for BPD Police Captain. As part of this examination, he submitted a Response Booklet containing approximately twelve single-spaced, handwritten pages of written answers addressing each of the three required essay questions of the In-Basket Test component. (*Exh. C17*)

12. Lt. Wilbanks received the following competency scores from the two assessors who graded each of the three exercises in the In-Basket Test:

<u>Assessor A</u>	<u>Written</u>	<u>Interpersonal</u>	<u>Analyzing/Deciding</u>	<u>Managing</u>
Exercise 1	7	5	3	4
Exercise 2	7	4	4	5
Exercise 3	8	6	5	6
Overall	7	5	4	5

<u>Assessor B</u>	<u>Written</u>	<u>Interpersonal</u>	<u>Analyzing/Deciding</u>	<u>Managing</u>
Exercise 1	8	7	5	4
Exercise 2	7	4	4	3
Exercise 3	7	4	5	4
Overall	7	5	5	4

(Exh. C16; Testimony of Director Ward)

13. In four cases, Assessor A circled more than one numerical score on Lt. Wilbanks' Consolidation Form. On Exercise 1, the assessor circled both an "8" and a "7" on Written Communications, an "8", a "4" and a "3" on Analyzing & Deciding and a "5" and a "4" on Managing Activities. On Exercise 2, the assessor circled both a "5" and a "4" on Analyzing & Deciding. In each case, the lowest number had the darker circle and was treated as the assessor's final scoring decision by EB Jacobs. Director Ward did not question these multiple markings or seek to confirm that the raw numerical scores EB Jacobs used in those instances were, in fact, the assessors' intended actual marks. (Exh. C16; Testimony of Director Ward)

14. In general, when two test assessors come up with different scores for the same candidate's test submission, it is considered "best practice" within the testing industry that the two assessors discuss their scoring and attempt to reach a consensus. EB Jacobs' specifically instructed the assessors that their scoring must be done independently and the two scores would be averaged. Although assessors would discuss their assessments for purposes of preparing a narrative "feedback" report on each candidate, "[n]o ratings will be modified once the feedback discussion begins". (Exhs C7 & C9; Testimony of Ward)

15. After "rescaling/weighting", Lt. Wilbanks received a Final Overall Examination Score of 89.72355, composed of the following examination component scores:

<u>EXAMINATION COMPONENT</u>	<u>SCORE</u>
Technical Knowledge	32.00000
In-Basket Test	15.90299
Oral Board Test	<u>22.44056</u>
FINAL EBJ EXAMINATION SCORE	70.34355

EDUCATION AND EXPERIENCE POINTS	17.38
VETERAN'S POINTS OR 25-YEAR SENIORITY POINTS	<u>2</u>
<u>FINAL OVERALL EXAMINATION SCORE</u>	<u>89.72355</u>

(Exh. C15)

16. Lt. Wilbanks' final overall examination score, after rounding to a whole number as provided by HRD Personnel Administration Rules, placed him on the current BPD Police Captain's Eligible List with a passing score of 90. (Exh. C15; Administrative Notice [PAR.07(4)])

17. On December 17, 2015, Lt. Wilbanks duly sought a review and recalculation of the scores he received on the In-Basket Test of his BPD Captain's examination. On December 16, 2015, he also sought a review of the Oral Board Test components of his BPD Captain's examination. (Exh. I)

18. After his requests were denied, Lt. Wilbanks appealed to the Commission which allowed his appeal in part, ordering that HRD conduct a further review of his In-Basket Test score and upholding HRD's determination that the Oral Board Test was not an "essay" test within the meaning of G.L.c.31,§24(a), dismissing that portion of his appeal for lack of jurisdiction. (Exhs. 1 & 2[Wilbanks I])⁷

19. As a result of the Commission's Decision in Wilbanks I, HRD's Director of Test Development, Briana Ward, was assigned to perform a review of the markings of Lt. Wilbanks' In-Basket Test. (Exh. 3; Testimony of Director Ward)

20. 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

⁷ The Commission's decision in Wilbanks I, which determined that the Oral Board Test was not an essay test, is pending judicial appeal.

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*)

21. 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 Lr. Wilbanks' 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 EB Jacobs provided to her, as well as her formal education and experience in Industrial & Organizational psychology. (*Testimony of Director Ward*)

22. Director Ward's review of the markings of Lt. Wilbanks' In-Basket Test began with the receipt of all the documents from EB Jacobs that Director Ward deemed necessary for her review. 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 each candidate received at the test center, the information associated with three scenarios to be reviewed and absorbed prior to making 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. Lt. Wilbanks’ In-Basket Response Booklet, Test Consolidation Form (assessor’s raw score sheet) and Notice & Feedback Report – Copy of Lt. Wilbanks’ written responses to the three In-Basket exercises, the assessors’ scores (1 to 9) assigned for each competency and overall score for each exercise, and a feedback report provided to Lt. Wilbanks containing “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. 3, C7 through C17; Testimony of Director Ward)

23. 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 Assessor's Training Manual described above. (She did not write down her scores but kept them "in her head.") She then compared her mentally assigned scores to the six candidates' raw scores assigned by the assessors. 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*)

24. Director Ward performed the same review of Lt. Wilbanks' Response Booklet. She mentally scored his responses and then compared her scores with the numerical raw scores assigned by the assessors. She testified that, in every case, her scores "aligned" with the assessors' numerical raw scores. (*Exh. C16; Testimony of Director Ward*)

25. Director Ward was not provided with the identity of the assessors or how many different assessors' panels were used. I infer, however, from the distinctly different styles of notetaking that appears on the assessors' Consolidation Forms, that:

- The same two assessors who evaluated Lt. Wilbanks (i.e., Assessor A & B) also comprised the panel that evaluated sample candidate "Mod 2", one of the sample "Moderately Effective" candidates.
- The panel that evaluated the other "Moderately Effective" sample candidate (Mod1) was comprised of Assessor A and one different assessor (i.e., Assessor C)

- All the “Highly Effective” and “Ineffective” sample candidates (High 1 and High 2, Low 1 and Low 2) were evaluated by a completely different panel comprised of two assessors (i.e., Assessors D & E), neither of whom were members of any panel that evaluated Lt. Wilbanks or either of the two sample candidates rated “Moderately Effective”.

(Exhs. C14 & C15; Testimony of Director Ward)

26. Director Ward also reviewed the excerpts from the Notice and Feedback Reports for the six sample candidates and for Lt. Wilbanks. She did not analyze the data they contained. *(Exhs. C14 & C15; Testimony of Director Ward)*

27. Also, the portion of the In-Basket Notice and Feedback Reports for the six sample candidates EB Jacobs provided to HRD contained the average raw assessors’ scores but did not include their “rescaled/weighted” scores. Therefore, Director Ward did not replicate or perform such analysis herself or specifically explain how the candidates’ “rescaled/weighted” scores were computed or how those candidates’ final scores compared to Lt. Wilbanks’ scores. *(Exhs. C14 & C15; Testimony of Director Ward)*

28. By letter dated February 24, 2016, Director Ward notified Lt. Wilbanks that she had completed her review of the markings of his In-Basket Test. The letter concluded: “I reviewed your In-Basket responses and 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 have concluded that you were scored accurately on the Police Captain In-Basket Test.” *(Exh. 3)*

29. This appeal duly ensued. *(Exh. 4)*

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, Lt. Wilbanks’ appeal challenges HRD’s review of the marks on his “In-Basket” Test scores, as the Commission directed in Wilbanks I.⁸ Thus, this appeal invokes that part of G.L.c.31, §22 through §24 which currently provides, in relevant part:

“ . . . [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

⁸ Similar issues are presented in Clarke II.

differed significantly from the current version of G.L.c.31,§22 through §24. See Wilbanks I.⁹ As this record showed, until this appeal, HRD’s incumbent 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. 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 decision that HRD is obliged to conduct an in-house substantive review of the markings of Lt. Wilbanks’ In-Basket Test, subject to appeal to the Commission. The parties differ, however, on the standard of review that must be applied

⁹ Lt. Wilbanks’ appeals are 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 or in Wilbanks I. 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)

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 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 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 thorough 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 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 Captain’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.

Mr. Wilbanks correctly cites Boston Police Sup. Officers Fed’n v. Civil Service Comm’n, 35 Mass.App.Ct. 688 (1993) for the proposition that the Commission “hearing” is intended to be a more formal proceeding than the “review” conducted by HRD, and that “the Legislature intended the commission acting in its quasi-judicial capacity, and not the administrator, to be the fact-finder relative to the fairness of the examination.” This opinion, however, must be read in light of the issue that was presented to the court, namely, a “fair test” appeal that challenged the fairness of the examination component that purportedly tested supervisory skills and the administrator’s impartiality in designing that component. These principles that govern a fair test appeal certainly bear notice, but the scope of an appeal from the marking of specific examination questions, focuses mostly on the answers given, rather than the fairness of the questions posed, so the analogy of the “fair test” line of cases is not precisely apt.

I have carefully considered the relevant legislative history and case law 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 conduct a “de novo” hearing to supesede

HRD's "review" of the grading of an essay question. In particular, I do not agree with Lt. Wilbanks' contention that on appeal from HRD's review of the marks of a single candidate's answers to an examination question, the Commission's "fact-finding" authority should go so far as require the Commission to put on "assessors' hats" and fix and establish ourselves, ab initio, the most "accurate" scores for Lt. Wilbanks' In-Basket Test responses. As a general rule, I do not find that would be appropriate or consistent with the Commission's appellate role. Lt. Wilbanks' has a point that the difference in "expertise" at HRD and at the Commission in the particular circumstances of this case may seem, to some, more theoretical than real, especially in a case such as presented here, when HRD did not design or administer the test in question. However, the statutory distinction in the powers and duties of each body must take primacy over any such perceived shortcomings of the individual incumbent officials from time to time.

Thus, it suffices that, in an appeal from the review of marking of an examination question, the Commission hew to its traditional quasi-judicial appellate oversight of other HRD "actions or inactions", namely, to ensure that HRD's decision is based on the type of "impartial and reasonably thorough" review that has been required of the Commission when it reviews other HRD "actions or inactions" under G.L.c.31, Section 2(b), and to ensure that HRD's conclusions are neither "arbitrary" nor "devoid of logic", but, rather, supported by a preponderance of credible evidence. That said, I did thoroughly examine the assessors' training materials, the In-Basket Test information booklet, the Response Booklets of Lt. Wilbanks and the six sample candidates and other materials provided by EB Jacobs for the purpose of confirming whether, in fact, by a preponderance of evidence, the materials were sufficient to permit HRD to make the required review and to support a conclusion by the Commission that HRD conducted its review

thoroughly and in a manner that reasonably replicated the assessors' process for scoring the In-Basket Test as HRD contended it had done.

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 proffered to support the conclusion that Lt. Wilbanks' In-Basket Test was correctly marked and to justify HRD's decision to decline to adjust it upward.

The Sufficiency of HRD's Review

Lt. Wilbanks does not dispute, and I find, that the materials that Director Ward examined prior to conducting her review of the In-Basket Test included all the information that reasonably appeared necessary and appropriate to HRD's review under the circumstances of this case.¹⁰ This preliminary review included the material Director Ward needed to obtain a general understanding of the subject matter covered by the In-Basket Test, as well as the components

¹⁰ Lt Wilbanks 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.

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.

Lt. Wilbanks does contend that Director Ward's application of the information she obtained from these materials is flawed in several respects. These flaws, he contends, require a finding that the review was not thorough and that Director Ward's conclusion must be set aside as arbitrary and capricious. Specifically, Lt. Wilbanks argues that Director Ward ignored significant scoring discrepancies in nearly half of the scores awarded by the two assessors to Lt. Wilbanks' answers. He also claims that Director Ward ignored indicia on the score sheets that allegedly showed that the assessors had violated certain protocols they were meant to follow in scoring the answers. Finally, he argues that Director Ward's contention that, when she made her own independent scoring of Lt. Wilbanks' answers, they "aligned" with the assessors' scores, is logically untenable and unsupported by the evidence.

In most respects, Lt. Wilbanks' arguments are not well-founded. The protocol for marking the In-Basket Test responses dictated that the assessor use a two-step decision-making process. First, the assessor must decide which one of three sections of the rating scale (highly effective, moderately effective or ineffective) applied to the candidate's response. Second, the assessor must decide on a specific rating point within the selected section based on the degree to which the candidate displayed the nine or ten specific characteristics (A-I for Written Communications and A-J for the other competencies) associated with the selected section of the scale.

Under this protocol, two point differences between assessors, or a difference in the general section assigned by the assessors, does not necessarily have consequences that Lt. Wilbanks suggests. I note that, each overall score assigned by each assessor for the four competencies – which is the score actually used to arrive at the “Overall In-Basket Raw Score” assigned to Lt. Wilbanks – consistently are within one point and within the same section of the rating scale:

<u>Competency</u>	<u>Assessor A</u>	<u>Assessor B</u>	<u>Average</u>
Written Communication	7	7	7.00000
Interpersonal Interactions	5	5	5.00000
Analyzing & Deciding	4	5	4.50000
Managing Activities	5	4	<u>4.50000</u>
OVERALL IN-BASKET RAW SCORE			<u>21.00000</u>

Even as to individual scores on the three exercises, no discrepancies raise per se red flags that would necessarily demand heightened scrutiny. In three two-point differences on specific exercises, Assessor A had given the higher score (Interpersonal, Exercise 3; Managing Activities, Exercises 2 & 3); in the other two instances, Assessor B had given the higher score (Interpersonal, Exercise 1; Analyzing & Deciding, Exercise 3). Thus, this suggests no pattern of being systematically “easy” or “hard” grading that might violate the scoring instructions.

Moreover, merely pointing to assessors’ scores assigned to different sections of the rating scale does not necessarily give cause for heightened scrutiny. For example, Lt. Wilbanks points to the two-point difference in the scores he received for Interpersonal Interaction on Exercise 1 to illustrate how Assessor B was “correct” (by awarding a “7” [the low end of the “Highly Effective” section]) and Assessor A did not “get that one right” (by awarding a “5” [the mid-point of the “Moderately Effective” section]).

Differences in individual exercise raw scores from “7” to “5” are not inherently as problematic as Lt. Wilbanks suggests. A “7” is a “borderline” score which means that, as to the particular answer, the assessor decided that that it met “about half “ of the ten evaluative criteria

(A-J) that fall within the Highly Effective section and “about half” in the Moderately Effective section, with there being “somewhat more in High than Moderate”. A “6” is also a “borderline” score which means that “about half “ of the ten evaluative criteria (A-J) that fall within the Highly Effective section and “about half” in the Moderately Effective section, with there being “somewhat more in Moderate than High.” A “5” is also a “borderline” score, which means that the answer met “most” of criteria in the Moderately Effective section, with the rest in another (higher or lower) section.

Also, the evaluative criteria for each of the three sections are not always easy to apply. For example, in evaluating Interpersonal Interactions, a candidate’s response was considered “Highly Effective” if it “demonstrated respect for the values and perspectives of others”; a candidate was “Moderately Effective” if he/she “showed some tolerance or respect for the values and perspective of others.” (In another example, Written Communications that “used highly effective wording to convey his/her thoughts” was considered a “Highly Effective” trait; using “words or phrases that were appropriate for the situation” was considered “Moderately Effective” behavior.) In sum, assuming that the assessors followed the assessment criteria in good faith, differences between Assessor A’s “5” and Assessor B’s “7” on Exercise 1, Interpersonal Interaction, falls well within a range of consistency in scoring that was not shown necessarily to be arbitrary or devoid of reason.¹¹

For similar reasons, I credit Director Ward’s testimony that her own assessment of Lt. Wilbanks’ answers “aligned” within two points of the assessors. While it would clearly have been preferable that Director Ward actually followed the training protocol by keeping notes and

¹¹ I do not discount the potential that, at some level, the ambiguity in distinctions between rating points could be found so overly subjective as to become illusory and to call into question whether the essay examination was a “fair test” of candidates’ relative abilities. As noted earlier, however, this appeal concerns the process for marking a candidate’s specific answer. It is not a fair test appeal. This concern, therefore, is not properly within the jurisdiction of the Commission for consideration in this appeal.

writing down her scores, her failure to do so here does not, alone, discredit her testimony. I accept that, at least as to the “overall” raw scores which are the crucial ones, her assessment was consistent with the scores reported as the “In-Basket Test Average Ability Scores” and the OVERALL IN-BASKET RAW SCORE” assigned to Lt. Wilbanks in his Score Notice and Feedback report.

I do concur with Lt. Wilbanks that what seem to be multiple markings on Lt. Wilbanks’ score sheet (and also on some of the sample test candidates’ score sheets) is problematic. The appearance of these markings do raise, above a speculative level, that the assessors may not have followed their instructions and may have consulted with each other before making a final score decision, may have given tentative assessments before completely reviewing the answers, or it may indicate that the assessor intended to award a point score between two whole numbers (i.e. marking “8” and “7” meant to award a score of “7.5”). However, after reviewing the assessors written comments on their score sheets and comparing the overall scores assigned, I am persuaded that these discrepancies, if any, did not materially impact the scores that were used to calculate Lt. Wilbanks’ overall scores as reported in his Score Notice and Feedback Report. Thus, while HRD’s review may be considered less than thorough in this regard, it does not rise to the level that would require any adjustments to Lt. Wilbanks’ marks.

Finally, I note that, even if all of the foregoing discrepancies that Lt. Wilbanks points out did, in the aggregate, call into question the thoroughness or arbitrariness of HRD’s review, any correction to his scores to which Lt. Wilbanks therefore would be entitled would make no material difference to his placement on the current BPD Captain’s Eligible List. Lt. Wilbanks currently sits on that list ranked with a whole number score of 90. After consideration of the entire record and all of the arguments presented, the largest increase that the Commission would

make to adjust for the discrepancies Lt. Wilbanks asserts to his Overall In-Basket scores, is as follows:

<u>Abilities</u>	EB Jacobs Test Average	
	<u>Ability Scores</u>	<u>Requested New Score</u>
Written Communication	7.00000	7.00000*
Interpersonal Interactions	5.00000	6.00000
Analyzing & Deciding	4.50000	5.00000
Managing Activities	<u>4.50000</u>	<u>5.00000</u>
OVERALL IN-BASKET RAW SCORE	21.00000	23.00000

*Lt. Wilbanks did not argue that the Commission should adjust his score for Written Communication (Appellant’s Proposed Decision, p. 34, fnt.13).

(This adjustment is actually higher than the score of 22.5 that Lt. Wilbanks presented in his Proposed Decision, Proposed Finding 47).

Using an adjusted OVERALL IN-BASKET RAW SCORE of 23.00000, Lt. Wilbanks In-Basket Rescaled/Weighted Score would be increased to 16.59710. This adjustment would increase his FINAL EBJ EXAMINATION SCORE to 71.03766 and, after adding 17.38 points for E&E and 2 Veteran’s/Seniority Points, his FINAL OVERALL EXAMINATION SCORE increases to 90.41766. When this new score is rounded, it results in the same whole number – 90 – as the whole number currently used to rank Lt. Wilbanks on the current BPD Captain Eligible List.

CONCLUSION

Accordingly, for the reasons stated, the appeal of the Appellant, Sean M. Wilbanks is *dismissed*.

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:

James F. Lamond, Esq. (for 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)(1), 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).