Audit Report

2015 Massachusetts Trial Court Probation Department Essay Examination

July 7, 2015

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In this report, Industrial/Organizational Solutions, Inc. (IOS) presents the results of an audit of the 2015 Massachusetts Trial Court (MTC), Probation department promotional essay examination rating process for the positions of Assistant Chief Probation Officer and Chief Probation Officer. The purpose of the audit was to evaluate the integrity of the rating process with regard to the accuracy and completeness of the raters' documentation, adherence to IOS procedures, occurrences of unethical behavior or breaches of confidentiality and anonymity, and the influence of rater assignment on the passing or failing of candidates.

IOS provided training to the raters in which they were instructed to follow very specific procedures. IOS interviewed each rater to evaluate their understanding of the rating process and any occurrences of unethical behavior. All raters accurately recalled the procedures outlined by IOS and did not report any unethical behavior in the rating process. All ratings were also provided in a "blind" process in which the raters did not know the candidate's name or other demographic information at any point.

Raters utilized rating criteria booklets (i.e., rubrics) that defined effective response behaviors as their primary documentation of their evaluations. Thus, the presence and accuracy of these booklets was paramount to the rating process. In total, eight MTC raters evaluated 298 candidates across up to three rounds of ratings, for a total of 504 rating criteria booklets. Ten booklets (1.98%) required verification of the rater or candidate ID numbers and were easily corrected. Within each of the 504 evaluations, raters assessed three separate responses, totaling 1,512 question-level evaluations. Thirteen instances occurred (0.86%) where the rater did not document any criteria being met by the candidate for a question, suggesting the potential for incomplete documentation. In each of these instances, the candidate received a score of 1 or 2, which on the rating scale indicates "poor performance; significant negative behaviors displayed". This indicates that the absence of criteria check marks resulted from the candidate failing to demonstrate positive performance criteria. Overall, raters adhered to the procedure outlined by IOS with regard to documenting performance using the rating criteria booklets.

The rating criteria also served as the basis on which raters provided 7-point ratings for each question; more criteria demonstrated in the responses should equate to a higher score. Indeed, strong positive (and statistically significant) correlations between the number of criteria achieved and the 7-point ratings were observed across all raters for each question from a random sample of 60 candidates (comprising 103 evaluations). Raters accurately assigned the 7-point ratings based on candidate performance relative to the pre-determined standard.

The notion that the rater(s) by which a candidate was evaluated determined whether or not they passed/failed the exam was unsubstantiated. IOS conducted statistical significance tests to evaluate the presence of this phenomenon and found no statistically significant pattern in the pass rates across all raters.

As a result of the audit, IOS concluded that the rating process for the 2015 Massachusetts Trial Court, Probation department promotional essay exam was fair, consistent, and a valid implementation of the procedures outlined by IOS. IOS also found the raters' documentation to be complete and did not uncover any problematic incidents in the process that threaten the validity of the exam. Without reservation, IOS recommends that the Massachusetts Trial Court Human Resources department use the results of the 2015 promotional examination to create eligibility lists for the above positions.

RATING PROCESS OVERVIEW

Industrial/Organizational Solutions, Inc. (IOS) performed an audit of the rating process to ensure that the Massachusetts Trial Court (MTC) raters adhered to the rules, procedures, and expectations outlined to them by IOS. The purpose of the following section of the report is to document and describe the rating process used to assess and score the essay section of the 2015 MTC Probation department promotional process for the Criminal Court Division. This section summarizes the rating process, scoring procedures, and cutoff score used in the process.

The Essay Exam

The essay exam consisted of three knowledge- and scenario-based questions that assessed candidates' knowledge, skills, and abilities (KSAs) required for the positions of Assistant Chief Probation Officer and Chief Probation Officer. The questions were designed collaboratively by IOS and MTC subject matter experts (SMEs) in the Probation department. Along with each question, IOS and MTC SMEs identified elements of an effective response, hereafter termed *rating criteria*. Candidates were given 60 minutes to complete all three responses and used "blue book" booklets to hand-write their responses. Candidates were only identified in the process by a unique ID number. All candidates were given the same questions, time limits, and materials, and were evaluated by the same set of rating criteria.

The Raters

The following eight SMEs from the MTC Probation department served as the raters for this process:

- Andy Peck
- Betsy Daigneault
- Dianne Fasano
- Francine Ryan
- Harriet Beasley
- Jeff Akers

- Pamerson Ifill
- Renee Payne

Training

Brian Marentette, Ph.D. from IOS provided a 3-hour training for all raters involved in the process. Raters were required to attend the training or watch the video-taped version of the training. The training was based on IOS' proprietary assessment method and covered the following broad topics:

- Rater Roles and Responsibilities
- The IOS Scoring System
- The IOS Rating Process
- Preventing Rating Errors and Biases
- Assessment Content Training (questions and rating criteria) and Practice

The Rating Process

Each essay was rated independently by a single rater. All essays were randomly assigned to raters and rated anonymously. Candidates were identified by a unique ID number. Corresponding names were known only to MTC HR staff and IOS. Raters were not provided any information related to the ID number (name, gender, ethnicity, etc.). Candidates' responses were evaluated against a predetermined set of rating criteria, which were observable statements reflecting positive behaviors that the candidate could display in their response. The rating criteria were developed jointly by IOS and the MTC SMEs to be consistent with the agency's source materials (e.g., Probation Standards), best practices, and minimum job expectations for written communication. The rating criteria were provided as guidelines for the raters to utilize as the objective standard against which each candidate could be evaluated, but they were not considered all-encompassing. In instances where candidates presented a behavior in their response that was effective, but was not listed in the rating criteria, the raters had the discretion to consider that in their evaluation. The same was true when candidates presented an ineffective behavior.

The raters utilized a step-by-step process to evaluate each candidate's responses against the set of rating criteria. First, raters read the essay response in its entirety. Raters took notes

or marked their observations of the behaviors displayed in the essay response on the rating criteria checklist. After the rater finished reading the response, they documented their observations on the rating criteria sheet. The rating criteria sheet then became the basis for assigning a rating on the 7-point rating scale. As a general rule, the more effective behaviors displayed in a candidate's response, the greater the 7-point rating. A very small degree of discretion was given to the raters so that they could utilize their subject matter expertise to provide more refined ratings. For example, a candidate may have demonstrated ten positive behaviors in the response, but failed to demonstrate two of the most critical behaviors. This response would not be rated as highly as another candidate who also demonstrated ten positive behaviors, but demonstrated all of the most critical behaviors. All raters were subject matter experts in this content area and were qualified to make this type of distinction. Thus, raters were given this small degree of discretion. As mentioned previously, in all cases the 7-point rating was based largely on the number of rating criteria demonstrated in the candidate's response with some consideration for the holistic effectiveness of the entire response. Raters determined the rating on the 7-point scale based on their observations for each question. The raters then documented these ratings on a final rating form. The raters repeated this process for each of the three essay responses.

Pass/Fail (cut score) Determination

The cut score was determined using industry best practices and IOS' standard procedure for evaluating this type of assessment. The midpoint (a rating of 4) on the rating scale was defined as a "Minimally Qualified" response. In this assessment, minimally qualified was defined as achieving approximately 40% of the rating criteria, with consideration for the overall effectiveness of the response and other behaviors displayed. Candidates were required to score, on average, a "4" for each response to pass the essay exam (i.e., a sum of 12 for the three questions).

Failure Procedure

Candidates who scored at or above the minimally qualified point for the sum of their three responses were deemed to have passed and were not subjected to further evaluation. Candidates who scored below the minimally qualified point were deemed to have failed and

were subjected to a second evaluation. A second independent evaluation was provided by a different rater using the same process described above. These raters were not provided the previous rater's evaluation. If the second evaluation resulted in another failure, the candidate failed the exam. If the second evaluation resulted in a pass, the candidate was subjected to a third independent evaluation by a different rater using the same process describe above. Again, these raters were not provided the scores from the previous raters during the rating process. If this third evaluation resulted in a pass, the candidate passed the exam. If the third evaluation resulted in a fail, the candidate failed the exam.

Scoring Procedure

After all candidates had been evaluated, the final rating sheets were sent to IOS. IOS utilized their standard scoring procedure in which the rating sheets were scanned in, checked for accuracy, then translated into data files containing the candidate ID numbers, rater ID numbers, and scores for each question. IOS consultants computed the final scores and merged the candidate information (name, position tested, etc.) with the rest of the essay data. IOS verified the number of candidates, ID-name matches, and other key information with MTC HR to ensure accuracy and completeness. IOS prepared the final results and delivered them to MTC HR after computing all final scores and conducting extensive quality control processes at IOS.

AUDIT RESULTS

As outlined above, IOS trained MTC raters to utilize a very specific process for evaluating the essay responses. The procedures outlined by IOS are designed to maximize accuracy, objectivity, anonymity, and fairness in the evaluation process. IOS thoroughly examined each rater's account of the process, documentation, and resulting scores to ensure that MTC raters conformed to the procedures required by IOS. IOS consultants also inquired with each rater about any impropriety or other unethical behavior that occurred during the process through one-on-one interviews.

Final Rating Sheets

The final rating sheets underwent an extensive quality control and verification process as part of the initial scoring prior to this audit. As part of the initial scoring process, IOS corrected stray marks, corrected ID numbers, and addressed other administrative issues that could be easily resolved. A second review of the sheets was conducted during this audit. During the course of the second review by IOS, we identified one candidate's final rating sheet that was added incorrectly, resulting in a 'pass', when it should have been a 'fail'. This candidate should have received a second (or potentially third) evaluation but did not. Thus, this candidate was rated again using the same procedures outlined by IOS. The raters' final documentation for these additional reviews was correct. The remaining 297 final rating sheets were correct with respect to the candidate ID numbers, rater ID numbers, final scores, and pass/fail status.

Use of Rating Criteria Booklets

In each rating booklet, raters were required to document the candidate ID number and their rater ID number, and mark the criteria that were demonstrated in the candidate's response. There were 298 evaluations in round one, 163 evaluations in round two, and 43 evaluations in round three (a total of 504 evaluations across all three rounds). IOS examined each of the 504 rating criteria booklets produced by the raters for adherence with the IOS documentation requirements. Two rating criteria booklets were missing the candidate and rater ID numbers. IOS was able to work with the MTC to identify the rater and the candidate to

which the booklets belonged. After working with MTC to identify the two incomplete booklets and complete the missing information, 100% of the booklets were accounted for. Below are the fields of data that raters were asked to complete for each evaluation, along with the corresponding result of the audit.

- Rater ID Numbers: 100% complete. Three booklets required verification to confirm the rater ID number during the audit, but were completed/corrected as a result.
- Candidate ID Numbers: 100% complete. Seven booklets required verification to confirm the candidate ID number during the audit, but were completed/corrected as a result.
- Question 1 Criteria: 100% complete. It was noted that six rating criteria booklets did not contain any checkmarks for Question 1. The scores for these six instances of Question 1 were 1, 2, 1, 1, 1, and 2. According to the rating scale, scores in this range indicate that the candidate demonstrated less than 10% of the rating criteria behaviors for that question; thus, the lack of check marks was indicative of the candidate not demonstrating any/enough positive behaviors.
- Question 2 Criteria: 100% complete. It was noted that four rating criteria booklets did not contain any checkmarks for Question 2. The scores for these four instances of Question 2 were 1, 2, 2, and 2. According to the rating scale, scores in this range indicate that the candidate demonstrated less than 10% of the rating criteria behaviors for that question; thus, the lack of check marks was indicative of the candidate not demonstrating any/enough positive behaviors.
- Question 3 Criteria: 100% complete. It was noted that three rating criteria booklets did not contain any checkmarks for Question 3. The scores for these four instances of Question 3 were 1, 2, and 1. According to the rating scale, scores in this range indicate that the candidate demonstrated less than 10% of the rating criteria behaviors for that question; thus, the lack of check marks was indicative of the candidate not demonstrating any/enough positive behaviors.

Rater Account of Rating Procedures

IOS Consultant, Linda Reynaud, M.A. interviewed each rater individually using the protocol presented in Appendix A. The purpose of the interviews was to determine, for each rater, if they attended the mandatory training, understood their responsibilities, felt that they could be consistent, maintained candidate anonymity, did not view previous round's evaluations, were not subject to any unethical influences, and followed the step-by-step rating process outlined by IOS. Below is a summary of the outcome of the interviews:

- Training: Raters were required to attend the training provided by IOS on the rating process. IOS viewed the training as critical and mandatory for the raters.
  - 100% of the raters received training. Six raters attended the training session in person, and two raters watched the full 3-hour video recorded version.
- Training Effectiveness: IOS sought to determine if the raters felt that they understood the training and were prepared to serve as raters in this process.
  - 100% of the raters felt the training provided them with the information they needed to rate the exams. 100% of the raters were able to have their questions about the process answered and felt prepared to serve as a rater in the process.
- Consistency in Rating: The IOS rating process was designed to maximize the fairness and consistency between raters and across candidates. IOS inquired about whether or not raters felt that the process required by IOS allowed them to be consistent in their evaluations across all of the candidates.
  - 100% of the raters felt the process required by IOS allowed them to make consistent ratings across candidates.
- Rating Process Steps (recalled by raters): IOS outlined a very specific step-by-step process for the raters to follow in their evaluations during the training. As a follow-up to determine if the raters had internalized this process and utilized it during the evaluations, IOS asked the raters to explain to IOS what process they used to make their ratings.
  - 100% of the raters outlined the steps in the rating process in a satisfactory manner when asked to explain to IOS what they did during their evaluation.

- <u>Previous Round's Ratings (if rating Round 2 or 3)</u>: As raters completed the Round 2 and Round 3 ratings, IOS asked MTC HR to keep the final ratings and booklets from the previous rounds out of view of the raters so as not to influence them during their subsequent ratings. Ultimately, because a single rating sheet was used to document all of the final ratings, the raters were able to see the previous ratings provided for each candidate when they documented their own final ratings. IOS was more concerned with the raters' ability to view the previous raters' rating criteria booklets and final rating sheets *before* or *during* their rating process.
    - None of the raters indicated that they were provided the previous raters' rating criteria booklets during their evaluations. Six of the raters indicated they were not able to view the previous round's final ratings. One rater stated that he was not able to recall. One rater stated that he was able to view the previous round's ratings, but only after he had finalized his ratings and was placing his final ratings on the final rating sheet (as expected). IOS concluded that there were no incidents of concern regarding raters viewing the previous ratings.
- <u>Unethical Influence on Raters</u>: IOS sought to determine whether the raters were asked to change their ratings or rate certain candidates a particular way, or if they were otherwise influenced by any type of unethical behavior from other MTC personnel (or anyone else).
    - 100% of the raters reported that they were not influenced by any members of the Trial Court (or anyone else) to falsify or otherwise provide different ratings than what they intended to provide.
- <u>Observation of Rater Impropriety</u>: Raters were also asked to disclose whether they observed any of their fellow raters engaging in unethical behavior during the evaluation process.
    - 100% of the raters reported they did not observe any unethical behavior or impropriety.

Utilization of the Rating Scale

The rating process for each question resulted in a single score on a 7-point scale. The 7-point rating was largely determined by the number of rating criteria demonstrated by the candidate in their response. A small degree of discretion was encouraged for instances where a candidate demonstrated or failed to demonstrate the most important rating criteria, or when a candidate presented a particularly positive or negative behavior in their response. Using this method, the number of criteria achieved should be positively related to the 7-point rating for that question.

IOS randomly sampled 60 candidates with representation from all eight raters. Across these 60 candidates, there were 60 first round ratings, 31 second round ratings, and 12 third round ratings for a total of 103 ratings. For each evaluation, IOS documented the number of criteria documented by the rater for each question, as well as the 7-point rating provided by the rater. IOS then computed bivariate correlations to examine the strength, direction, and statistical significance of the relationship between the number of demonstrated criteria and 7-point scores. According to the convention for interpreting a correlation coefficient provided by Cohen (1988), a correlation of .10 is small; a correlation of .30 is medium; and a correlation of .50 is large. Table 1 below displays the correlations between criteria demonstrated and the 7-point rating by question. The correlations are all considered large by convention and are statistically significant, indicating that there is a very strong and positive relationship between the number of criteria demonstrated by the candidate and their score on the 7-point scale. This provides strong evidence that the raters adhered to the rating process and utilized the 7-point scale consistently with IOS requirements.

Table 1: Correlations between number of demonstrated criteria and 7-point ratings by question

Question	r
1	.71**
2	.59**
3	.71**

** $p < .01$

Quantitative Review for Rater Scaler Bias

Human raters have the potential to mis-use a rating scale when providing their evaluations of performance. IOS analyzed the ratings for evidence of a severity/leniency bias, in which a rater may only use the top end of the scale (leniency) or the bottom end of the scale (severity). The potential result of this bias is a higher or lower pass rate for that rater compared to the other raters. A chi-square test of rater by pass/fail was conducted to evaluate the presence of this bias at a statistically significant level. The result of this test was non-significant (χ^2 [7] = 12.94, p > .05), indicating that the rater assigned to a candidate did not significantly influence the outcome of the candidate passing/failing the essay exam.

Conclusions

The purpose of the audit was to document the rating process and evaluate the integrity of the process in terms of the accuracy and completeness of the rater's documentation, adherence to IOS procedures, occurrence of unethical behavior, and presence of rater differences in pass rates. Based on the findings of the audit, IOS concluded the following:

- All raters accurately recalled the procedures outlined by IOS and did not report any unethical behavior in the rating process. The raters followed the procedures required of them by IOS.

- All ratings were provided in a "blind" process in which the raters did not know the candidate's name or other demographic information.

- The raters adhered to the procedure outlined by IOS with regard to documenting performance using the objective rating criteria.

- Raters accurately assigned the 7-point ratings based on candidate performance relative to the pre-determined standard. Strong positive (and statistically significant) correlations between the number of marked rating criteria behaviors and the 7-point ratings were observed across all raters for each question.

- The relationship between raters assigned to a candidate and that candidate's pass/fail status was non-significant, indicating an absence of any rater severity/leniency bias.

Based on the results of the audit summarized herein, IOS concluded that the rating process for the 2015 Massachusetts Trial Court, Probation department promotional essay exam was a valid implementation of the procedure outlined by IOS. IOS did not uncover any incomplete/missing documentation that could not be corrected, biases in the scoring, or other areas of concern or problematic incidents in the process. Without reservation, IOS recommends that the Massachusetts Trial Court Human Resources department use the results of the 2015 promotional examination to create eligibility lists for the positions of Assistant Chief Probation Officer and Chief Probation Officer.

CONTACT INFORMATION

Project Director:

Brian J. Marentette, Ph.D.
Manager, Consulting Services
IOS, Inc.
1127 S. Mannheim Rd., Suite 203
Westchester, IL 60154
888-784-1290

Consultant:

Linda Reynaud, M.A.
Consultant
IOS, Inc.
1127 S. Mannheim Rd., Suite 203
Westchester, IL 60154
888-784-1290

APPENDIX A – Audit Interview Protocol

IOS Rater Audit Interview Protocol

1. Did you attend the training provided by IOS or watch the video-recorded version?

    a. If no, please explain.

2. Do you feel the training gave you the information you needed to be a rater in this process?

    a. If no, please explain.

3. After all questions were answered, did you feel prepared to serve as a rater for the essay exams?

    a. If no, please explain.

4. Did you feel that the process you were instructed to use allowed you to make consistent ratings across all the candidates you evaluated?

    a. If no, please explain.

5. Please describe the steps you took in the rating of a booklet from beginning to end.

6. How did you use the rating criteria (pink sheets) during your evaluations of the booklets?

7. Were you ever aware of the candidate's name that you were evaluating?

    a. If yes, please explain.

8. When completing "round 2" or "round 3" ratings, did you see the previous rounds' ratings (either before or during your rating process)?

    a. If yes, please explain.

9. Were you influenced by any other members of the Trial Court (or anyone else) to falsify or otherwise provide different ratings than what you intended to provide?

    a. If yes, please explain.

10. Did you observe any impropriety in your fellow raters during the evaluation process?

    a. If yes, please explain.

11. If you had any questions, were you able to have them answered by HR or IOS?

    a. If no, please explain.