

The PBA vs Piscataway: A Case Study Statistics In The Workplace

Sheldon Epstein, Seton Hall University, USA
Aliza Rotenstein, Yeshiva University, USA
Bernard Dickman, Hofstra University, USA
Yonah Wilamowsky, Seton Hall University, USA

ABSTRACT

The State of New Jersey Public Relations Employment Commission recently rendered a decision in a dispute over the Piscataway Police Department's procedure for promoting individuals to the rank of sergeant. One important component of the case was how to properly interpret the results of a 1999 sergeant's promotion exam. This paper gives a brief history of the promotional process and offers the data and statistical analysis submitted by both the Plaintiff (Patrolmen's Benevolent Association- PBA) and the Police Administration. The case provides an excellent tutorial for beginners and practitioners on how to properly apply some elementary, but powerful, statistical concepts.

Keywords: Case Study, Statistics, Correlation, Rank

CASE BACKGROUND

In 1999, promotion to the position of sergeant in the Piscataway Police Department was based on a written test, an oral interview and seniority. The written test was given first and was conducted by an outside agency chosen by the Police Administration. Only applicants passing the written exam with a grade of 70 or more (out of a possible 100 points) were permitted to continue on to the oral interview. The oral interview was conducted and scored in-house by the Piscataway Police Administration and each interviewee was assigned a grade between 0 and 100. The written and oral exam scores were then weighted at 50% each and added to the applicant's number of years of seniority (up to a maximum of 10) to yield a final grade. The anticipated available sergeant positions were to be filled from the applicants list in descending order of final scores.

PLAINTIFF'S STATISTICAL CASE

Table 1 gives the actual results, in descending order of scores, from the 1999 sergeant's exam, in which eight positions were available and 23 applicants passed the written exam. The Piscataway PBA alleged that the oral test scores were manipulated by the Piscataway Police Administration, knowing the results of the written exam, in order to ensure a final ordering of candidates that was to the Administration's liking.

To demonstrate the "manipulation" of the numbers, the Plaintiff expanded Table 1 by adding three columns, ranking all 23 applicants who passed the written test, in each of the individual areas of written scores, oral scores and seniority. Table 2 lists these ranking results. A brief filed by the PBA then asserted: (see Table 1)

"The rankings are separated into 2 groups of eight candidates and a group of 7 candidates. The first group of 8 candidates represents the initial individuals who were promoted. What is clear by simply "eyeballing" the results is that there is a correlative relationship between the "Oral" ranking ... and the "Final" ranking... The "Oral" ranking, which had a weighted value of 50%, was within one to three points of the final promotional outcomes. Statistically revealing is that, for the most part, the other groups also had a correlation between the oral ranking and the respective candidates' final ranking."

Table 1: 1999 Sergeant’s Test Results

Applicant	Written Score	Weight 0.50	Oral Interview	Weight 0.50	Seniority	Final Score
1	87	43.5	90.2	45.1	10	98.60
2	86	43.0	85.0	42.5	10	95.50
3	88	44.0	81.6	40.8	10	94.80
4	72	36.0	91.2	45.6	10	91.60
5	79	39.5	84.2	42.1	10	91.60
6	87	43.5	77.2	38.6	7.75	89.85
7	78	39.0	80.4	40.2	10	89.20
8	80	40.0	75.0	37.5	10	87.50
9	90	45.0	64.0	32.0	10	87.00
10	80	40.0	70.0	35.0	10	85.00
11	82	41.0	68.0	34.0	10	85.00
12	84	42.0	65.0	32.5	9.75	84.25
13	78	39.0	68.6	34.3	10	83.30
14	84	42.0	60.8	30.4	10	82.40
15	72	36.0	79.2	39.6	6.5	82.10
16	74	37.0	68.0	34.0	10	81.00
17	87	43.5	53.2	26.6	9.75	79.85
18	81	40.5	57.4	28.7	10	79.20
19	83	41.5	55.2	27.6	10	79.10
20	86	43.0	51.8	25.9	10	78.90
21	74	37.0	65.2	32.6	6.5	76.10
22	72	36.0	58.4	29.2	10	75.20
23	74	37.0	53.0	26.5	10	73.50

The Plaintiff pointed out that a similar relationship between final ranking and written ranking is absent and concluded:

“One can aver that how well a candidate scores on the “Oral” category is significantly more important in determining their “Final” rank. This phenomenon reeks of manipulation and impropriety given the fact that both components are weighted at 50%.”

Table 2: 1999 Sergeant’s Test Results By Ranking

Candidate	Written Rank	Oral Rank	Seniority Rank	Final Rank
1	3	2	1	1
2	6	3	1	2
3	2	5	1	3
4	21	1	1	4
5	15	4	1	5
6	3	8	21	6
7	16	6	1	7
8	13	9	1	8
9	1	16	1	9
10	13	10	1	10
11	11	13	1	11
12	8	15	19	12
13	16	11	1	13
14	8	17	1	14
15	21	7	22	15
16	18	12	1	16
17	3	21	19	17
18	12	19	1	18
19	10	20	1	19
20	6	23	1	20
21	18	14	22	21
22	21	18	1	22
23	18	22	1	23

To quantitatively buttress the “eyeballing” argument, the PBA offered Table 3, which lists Coefficients of Correlation (CC) for each of the different segments of data previously identified:

Table 3: Correlation Coefficients Using Final Rankings

Candidates	With Respect To Oral Rank	With Respect to Written Rank
1 through 8	0.79	0.47
9 through 16	-0.31	0.72
17 through 23	-0.13	0.81

Using these numbers, the Plaintiff concluded:

- The CC between the ranking of the first eight candidates, with respect to the oral test and their final ranking, is 0.79, while the CC between the ranking of the same eight candidates, with respect to the written test and their final ranking, is a lower 0.47. This demonstrates that, although both the oral and written exams were ostensibly given the same weight of 0.5, the oral exam had a considerably greater effect in determining who was being promoted to sergeant.
- The CCs between the rankings of the second and third groups of candidates, with respect to the oral test and their final rankings, are somewhat negative, while the CCs between the rankings of the same two groups of candidates, with respect to the written test and their final rankings, are highly positive.

Table 4: Final Ranking Using Written, Oral & Seniority Ranks

1999 Sergeant's Test Results by Rank							
Applicant	Written Rank	Weight 0.50	Oral Rank	Weight 0.50	Seniority Rank	Ranking Score	Final Ranking
1	3	1.5	2	1.0	1	3.50	1
2	6	3.0	3	1.5	1	5.50	3
3	2	1.0	5	2.5	1	4.50	2
4	21	10.5	1	0.5	1	12.00	6
5	15	7.5	4	2.0	1	10.50	5
6	3	1.5	8	4.0	21	26.50	19
7	16	8.0	6	3.0	1	12.00	6
8	13	6.5	9	4.5	1	12.00	6
9	1	0.5	16	8.0	1	9.50	4
10	13	6.5	10	5.0	1	12.50	9
11	11	5.5	13	6.5	1	13.00	10
12	8	4.0	15	7.5	19	30.50	20
13	16	8.0	11	5.5	1	14.50	12
14	8	4.0	17	8.5	1	13.50	11
15	21	10.5	7	3.5	22	36.00	22
16	18	9.0	12	6.0	1	16.00	14
17	3	1.5	21	10.5	19	31.00	21
18	12	6.0	19	9.5	1	16.50	16
19	10	5.0	20	10.0	1	16.00	14
20	6	3.0	23	11.5	1	15.50	13
21	18	9.0	14	7.0	22	38.00	23
22	21	10.5	18	9.0	1	20.50	17
23	18	9.0	22	11.0	1	21.00	18

DEFENDANT’S PERSPECTIVE

The Defendant rejected the PBA’s methodology and inferences on the following grounds:

- In presenting its case, the PBA offered no motivation for changing the results of the different examinations from actual test scores to rankings. In fact, changing from numerical scores to rankings loses significant information. For example, with respect to Seniority (Table 2), Applicants 1 and 12 are ranked 1st and 19th, respectively, although their Seniority is almost identical (i.e. 10 vs. 9.75 years). Conversely, Applicants 12 and 6 are ranked 19th and 21st, but are actually much further apart (i.e. 9.75 vs. 7.75). Converting numbers

to ranks changes the dynamics of the system. Table 4 easily demonstrates this by applying the weighting scheme of Table 1 to the rankings, rather than the numerical scores, to calculate an aggregate value for each candidate:

Note that the final ordering in this situation (i.e. based on weighted rank in ascending order) is not the same as the one derived in Table 1. Applicant 12, for example, dropped from 12th to 20th, primarily based on a ranking of 19 for Seniority, even though he lagged in this category from the top candidate by only 0.25 units of service. When using the actual numbers, the 0.25 difference is diluted and plays only a minimal role. In a statistical analysis, the use of actual numbers is always preferable to the use of rankings. The former allows us to gauge how much better one candidate is than another. The latter does not.

- The correlation formula used to develop Table 3 is meant for actual numerical data, not for ranked data. Table 5 gives the corrected CCs based on the numerical data.

Table 5: Correlation Coefficients Using Numerical Values 3 Groups

Candidates	With Respect to Oral Score	With Respect to Written Score
1 through 8	0.69	0.53
9 through 16	-0.29	0.79
17 through 23	-0.26	0.90

Note the 0.79 CC for oral ranking, with respect to final ranking, has been reduced to 0.69 and the 0.47 for written ranking has been increased to 0.53 so that the 0.32 difference in CCs has been cut in half to 0.16.

- The PBA offered no rationale for decomposing the 23 Applicants into 3 groups (8, 8 and 7 respectively) rather than 2 groups of 8 and 15 respectively (i.e. those promoted and those not promoted). In conducting a statistical analysis, large groups of data are preferable to smaller ones. This case perfectly illustrates the point. Had the analysis been done with only two groups, the CCs for Candidates 9 through 23 in Table 3 would appear as seen in Table 6:

Table 6: Correlation Coefficients Using Numerical Values 2 Groups

Candidates	With Respect To Oral Score	With Respect To Written Score
1 through 8	0.69	0.53
9 through 23	0.56	0.52

Note that combining 9 through 23 into a single group, changes the two negative values for the oral tests in Table 5 to a single positive value of +0.56 in Table 6. In addition, the CCs for the written tests have dropped sharply. In fact, the CC for oral only slightly exceeds that for written among the applicants who were not promoted. Again, the volatility of CCs with respect to the amount of data used (i.e. the switching of CCs from negative to positive values as the number of items in the data set goes from 7 and 8 to 15) clearly illustrates why small subsets of data should not be used. In this case, when all of the data is analyzed together, the differences that allegedly point to “manipulation” disappear.

- The PBA’s underlying assumption that *it “reeks of discrimination” if it can be shown that the score on the oral test is more important in determining which candidates finish at the top of the list* is wrong. Assuming the written and oral tests are both administered fairly and are weighted at 50% each, it can easily be shown that the more important factor in determining the people who finish at the top of the list of candidates will appear to be the oral exam. The key to this analysis is that only candidates receiving a grade of at least 70 on the written exam are eligible for promotion. The same is not true for the oral exam; i.e. candidates failing the oral test (score under 70) are still eligible for promotion. Table 7 lists the Means, Standard Deviations and Ranges for each column of Table 1.

Table 7: Means, Standard Deviations, Ranges

	Written	Oral	Seniority	Final Score
Mean	80.8	69.7	9.6	84.8
Standard Deviation	5.7	12.2	1.1	6.8
Range	72 to 90	51.8 to 91.2	6.5 to 10	

For example, in Table 7, the Mean Seniority for all 23 candidates is 9.6, with a SD of 1.1. These values reflect the fact that the overwhelming majority of the candidates had Seniority values only a little higher or lower than 9.6. For the written exam, the Mean for all of the candidates is 80.8, the SD is 5.7 and the grades Range from a low of 72 to a high of 90 (i.e. 18 points). While this set of numbers is not as compact as the Seniority numbers, the distance from the Mean grade to the highest or lowest grade is only about 9 points. In contrast, the oral exam grades are the most dispersed of all with a Mean Grade of 69.7, a SD of 12.2 and a Range of grades from 51.8 to 91.2 (i.e. 39.4 points). Since:

- The difference between the highest and lowest grades on the oral exam is so much greater than the difference between the highest and lowest grades on the written exam, and
- Everyone entered with almost the same Seniority. Anyone scoring at the upper end on the oral exam is guaranteed to do considerably better than anyone scoring at the lower end. For example, any candidate who outscores another by more than 18 points on the oral exam is almost assured of placing ahead of the other candidate overall, regardless of what grade either received on the written exam (i.e. since the spread between the highest and lowest written exam grade is only 18 points).

The lack of correlation noted by the PBA is, thus, a consequence of the exclusion of people who failed the oral exam. The fact that the written exam, in this circumstance, appears not to have a significant effect on promotion reflects a more generally understood (and by many misunderstood) result in statistics. Consider, for example, the following analogy given by the sociologist Steven Goldberg: What is the role that weight plays in the performance of NFL offensive tackles? The heaviest offensive tackle is not necessarily the best. Indeed, the correlation between weight and performance among NFL offensive tackles is probably quite low. But they all weigh more than 300 pounds.¹ In the same way, the fact that all applicants must have a score of at least 70 on the written test may result in a low correlation between written test scores and promotion. This does not imply that the written test score is not important in determining promotion.

- The consistency of the data and the reasonableness of the selection of the eight people who were actually promoted can be seen in Table 8. Table 8 differs from Table 1 in that all candidates who “failed” the oral exam (i.e. received a grade of under 70) were deleted. Note that of all of the original applicants, only 10 passed both the written and oral tests, and only two were not promoted: one had the lowest passing score on the written exam and the other had the lowest passing score on the oral exam.

Table 8: Candidates Who Passed The Written And Oral Exams

Weight	100%	50%	50%	Final
Candidates	Seniority	Written Score	Oral Interview	Score
1	10	87	90.2	98.60
2	10	86	85.0	95.50
3	10	88	81.6	94.80
4	10	72	91.2	91.60
5	10	79	84.2	91.60
6	7.75	87	77.2	89.85
7	10	78	80.4	89.20
8	10	80	75.0	87.50
10	10	80	70.0	85.00
15	6.5	72	79.2	82.10

¹ See “Jewish Genius” by Charles Murray in *Commentary Magazine*, April 2007.

Finally, recalculating the CCs of Table 3 for all 10 candidates who passed both the written and oral exams yields Table 9:

Table 9: Correlation Coefficients Using Final Numerical Values

Candidates	With Respect To Oral Score	With Respect To Written Score
10	0.71	0.65

Thus, the initial major CC discrepancy reported by the PBA for the written and oral exams has been reduced from 0.28 (0.79 to 0.47) to less than 0.06 (0.71 to 0.65).

CONCLUSION

Based on these and other considerations, the Arbitrator found in favor of the Defendant. Excerpts from the Arbitrator’s decision can be found in the Appendix.

AUTHOR INFORMATION

Sheldon Epstein is a professor of computing and decision sciences at the Stillman School of Business, Seton Hall University. He holds a PhD in operations research from New York University and has done extensive business consulting. His research has appeared in: *Computers and Operations Research, Naval Research Logistics, Journal of the Operational Research Society, American Journal of Mathematical and Management Sciences, Opsearch, Annals of the Society of Logistics Engineers, Journal of Property Tax Assessment and Administration, Property Tax Journal, Location Sciences and Interface.*

Aliza Rotenstein is an assistant professor of accounting at the Sy Syms School of Business, Yeshiva University. She holds a PhD in business/accounting from the CUNY Graduate Center and teaches courses in financial and management accounting. Her main research interests include corporate governance, accounting restatements, and earnings management, and she is currently involved in research projects relating to accounting education and accounting for non-profit organizations

Bernard Dickman is an associate professor of information technology at the Zarb School of Business, Hofstra University. He holds a PhD in operations research from New York University and worked for many years as an operations research analyst for a major corporation. His research has appeared in *Computers and Operations Research, Naval Research Logistics, Journal of the Operational Research Society, American Journal of Mathematical and Management Sciences, OMEGA, The Mid-Atlantic Journal of Business and Location Sciences.*

Yonah Wilamowsky is a professor and former chairperson of computing and decision sciences at the Stillman School of Business, Seton Hall University. He holds a PhD in Operations Research from New York University. His research interests center on applications of statistics and quantitative methods to the law, business processes, biblical studies, and higher education. Among other journals, his work has appeared in *Naval Research Logistics, Journal of the Operational Research Society, American Journal of Mathematical and Management Sciences, Tradition, Journal of College Teaching and Learning, Property Tax Journal, Location Sciences and the Computers and Operations Research.*

APPENDIX

From Arbitrator's Decision:

The Arbitrator summarized the defendant's major arguments as follows:

- The PBA erroneously relied upon ranks, rather than numerical data, in conducting its analysis. This was inconsistent with the method used by the Township to calculate the candidate's final ranking, which was based upon the actual numerical results on each portion of the promotion process. Even more significantly, the PBA's reliance upon ranks, as opposed to the numerical data, resulted in the loss of valuable information and led the PBA to draw inaccurate conclusions when analyzing the data.
- The PBA's formula in computing Correlation Coefficients ("CCs") was statistically incorrect because it measured the correlation between rankings, rather than numerical data. When properly calculated using the numerical data, the difference in CCs with respect to the oral score significantly declined.
- The pool of data upon which the PBA relied was too small to ender statistically reliable results. This problem was exacerbated by the fact that, in calculating the CCs, the PBA randomly broke down the group of 15 candidates who were not promoted into two smaller subgroups.
- In calculating the CCs, the PBA, without explanation, broke down the group of 15 candidates who were not promoted into two seemingly random groups of eight and seven candidates. If these two subgroups are combined, it eliminates the negative correlation between the performance on the oral examination and the final score. In fact, when properly analyzed, the CCs for the oral test and the written are virtually identical (.56 vs. .52) indicating that each portion of the promotional process weighed equally upon the candidates final scores.
- The PBA's analysis is fatally flawed because it fails to take into consideration the fact that the numerical scores on the oral examination had a far greater range than the scores for the written examination, because the written examination had a minimum passing grade of 70, while the oral examination did not. The data analyzed by the PBA was limited to only those candidates who passed the written examination. The scores on the written examination thus ranged from a low of 72 to a high of 90, or an 18 point range. In sharp contrast, the scores on oral examination had a much wider range from a low of 51.8 to a high of 91.2, or a 39.4 point range. This difference fully and logically explains why the oral examination *appears* to play a more significant role on the final scores. If the oral examination scores are limited to a minimum grade of 70- as the written examination scores were limited- the mean and standard deviation for both examinations are virtually identical and the PBA's theory that the results on the oral examination played a greater role on the final rankings is wiped out.

NOTES