

Police Officer Selection Is Broken and We Can Fix It

Joel P. Wiesen, Ph.D.
October 15, 2020

ABSTRACT

Selection tests for police officer often result in two disturbing outcomes: (1) hiring of many new police officers who will fail on the job, and (2) hiring few minority candidates. Psychometric simulations reveal the reasons for and the extent of these disturbing findings. One administrative approach is described that is designed to help avoid one or both of these outcomes being replicated in real life.

This short article, the first of an anticipated series, can only begin to address a small part of this complex, important, and timely topic. I will address two very different facets: (1) Why police officer selection is broken, and (2) One easily implemented, administrative approach to help hiring managers avoid being surprised by the results of a police officer hiring process. In future articles I plan to present additional analyses of police officer selection and offer additional ideas on ways to both improve the expected job performance of police officers and hire a more diverse police officer workforce.

Police Officer Selection Is Broken

This article will describe two mathematical simulations of police officer hiring. (There are other reasons to think PO hiring is broken, but that is for later articles.) The mathematical simulations work like this. First, virtual applicant data is generated that reflects what our field often presumes is seen in the real world (e.g., setting test validity equal to that reported in published meta-analyses, and assuming a plausible number of candidates and openings). Then the simulation identifies the virtual candidates who would be hired based on their test scores. Finally, various aspects of the data for these hired candidates is summarized (e.g., to evaluate the level of adverse impact and expected job performance).

Simulation of the Use of a Multiple Choice Test of General Mental Ability

Consider this hiring scenario for a large jurisdiction: (a) corrected validity of general mental ability (g), $r = .24$; (b) 10,000 candidates, including 1,000 black candidates; (c) half of the job candidates can do the job; (d) a standardized B-W mean score difference (d) of 1 on the test of g ; and (e) a d of .24 for job performance (based on the regression formula $y = .24x$, where y is job performance, .24 is validity, and x is test score for g), as summarized in Table 1.

Table 1. Simulation Assumptions: Selection Using a Test of *g*

Validity: Test of *g*, $r = .24$

($r = .27$ for supervisor ratings, which becomes $.24$ when the correction for unreliability of the predictor is removed, Aamodt, 2004, e.g., page 35)

Number of candidates: 10,000 candidates, including 1,000 black candidates

Number of hires: Hire 1,000 new officers (top 10% of scores)

Quality of candidates: Half of candidates have the ability to be successful officers

B-W difference on *g*: $d = 1.0$

B-W difference on job: $d = .24$

This simulation shows that 33% of the hires are expected to fail on the job with only 13 black officers hired (rather than the expected 100). The expected adverse impact ratio is $.12$ (see Table 2), consistent with previous simulations (e.g., Wiesen, 2018).

Table 2. Simulation Outcomes for Selection Based on *g*

Hires who **cannot** do job: **329 (33%, chance = 50%)**

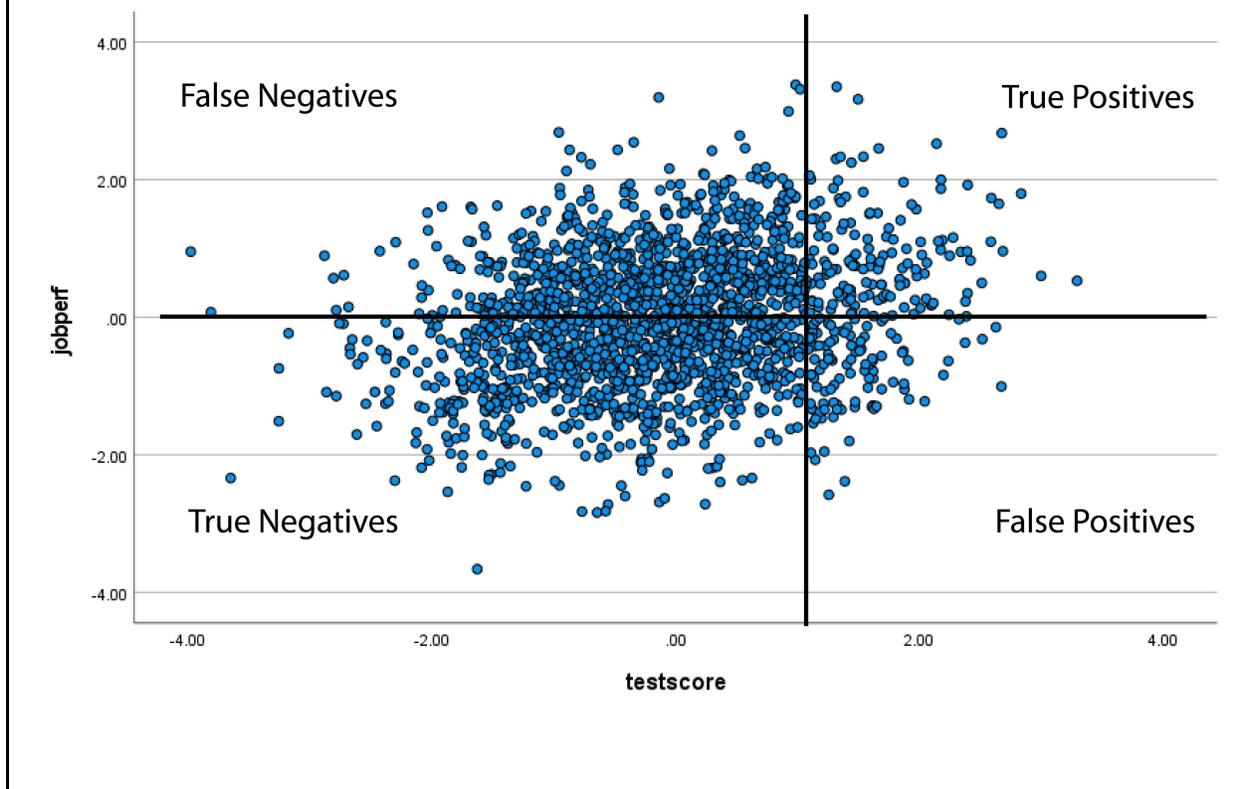
Black candidates selected: **13**

White candidates selected: **987**

Adverse impact: **.12** (severe)

A (reduced *n*) scatter plot shows a considerable number of false positives (in the lower right quadrant of Figure 1).

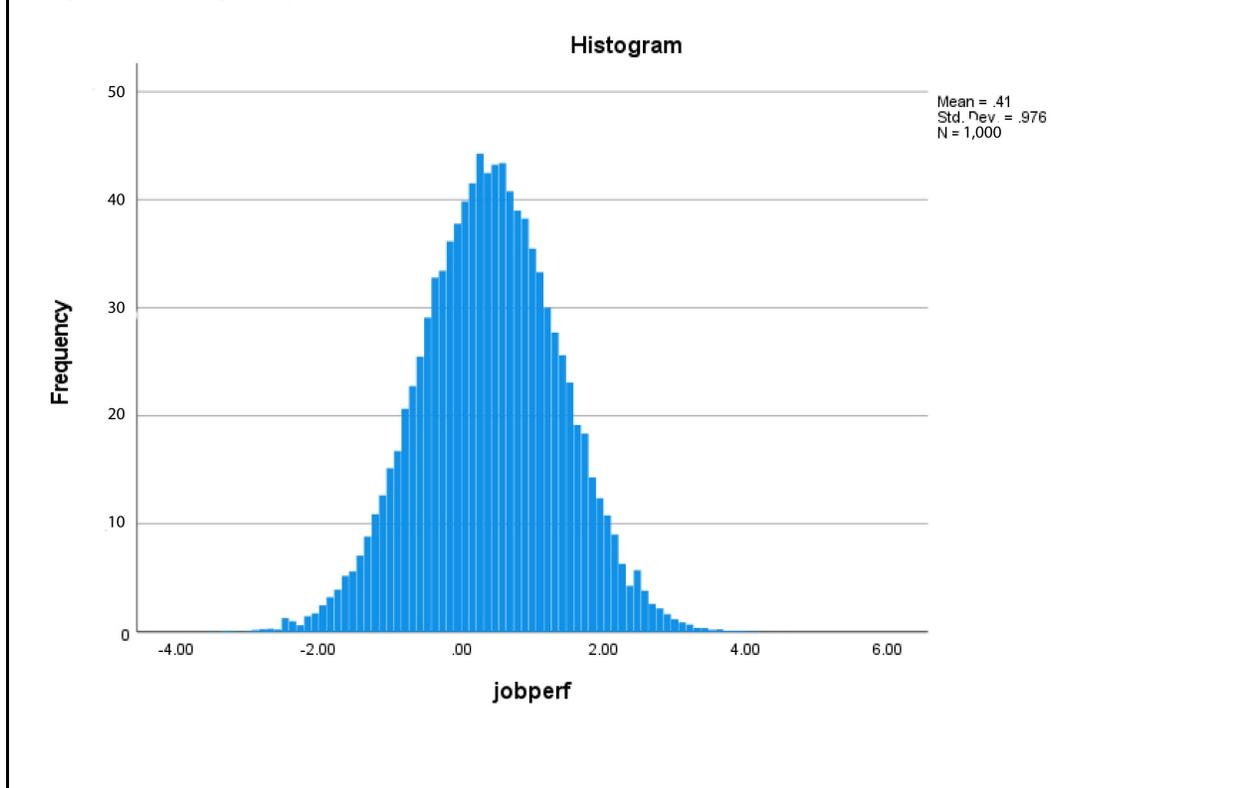
Figure 1. Scatterplot of Test Score and Job Performance Simulation



The proportion of new hires expected to be able to perform the job successfully can be expressed as an accuracy statistic known as the Positive Predictive Value (PPV), in this case .67. The PPV is used in other fields (e.g., medicine) as one of several prediction accuracy statistics. PPV is rarely used by IO psychologists; we prefer to use r . Hiring managers want all new hires to be capable of job success. Thus, hiring managers envision the world in terms of PPV rather than r . I recommend that the profession of IO psychology develop guidelines for acceptable levels of PPV. PPV is partly a function of the quality of the candidates (a topic for future articles).

Mean expected job performance reveals a similar disturbing pattern of poor job performance. The simulation results in 7 hires that are 2 or more standard deviations below the candidate mean in job performance, 74 hires that are 1 or more standard deviations below the mean, and 337 hires at or below the mean. (See Figure 2.) The mean job performance among those hired is .41, a substantial improvement over the z of 0 that would result from chance selection. Yet, 74 of those hired are well below the mean of all **candidates** in job performance.

Figure 2. Frequency Distribution of Job Performance for Those Hired



I take the results of this simulation as renewed impetus for searching for better ways to select police officers. In fact, the actual PPV of police hiring is probably below 50% based on the following logic. Police officers need to possess characteristics beyond g . Surely some of the non-measured, non- g characteristics are crucial and some job candidates lack those characteristics. If we assume that 50% of the candidates lack sufficient levels of these other characteristics to perform the job, then 50% of the 67% of new hires who were identified by the written test as true positives actually are expected to fail on the job. This leaves us with a PPV of 34%, meaning 2/3 of the new hires are expected to fail on the job. I take the results of this extension of the simulation as even greater impetus for searching for better ways to select new police officers.

Simulation of the Use of a Composite of g and Personality

Composites of measures of g and personality are (increasingly) used rather than measures of g alone. Such composites also are expected to result in high numbers of false positive hires and severe adverse impact, although somewhat less severe than hiring based on g alone, as shown by the next simulation.

Let us modify the simulation above as follows: (a) select based on a composite of g and a test of personality, (b) assume $r = .25$ for the personality test, (c) equally weight g and personality, and (d) assume B-W d of zero on the personality test (see Table 3).

Table 3. Additional Simulation Assumptions, Including a Personality Test

Validity: Personality test, $r = .25$

B-W difference: $d = 0$ for personality

Composite: Equally weight g and personality

The outcome of this modified simulation is that 26% of the hires are expected to fail on the job and the expected adverse impact ratio is .22 (see Table 4).

Table 4. Simulation Outcomes for Selection Based on g and Personality

Hires who **cannot** do job: **257 (26%)**

Black candidates selected: **25**

White candidates selected: **975**

Adverse impact: **.22** (severe)

The high number of false positives in this modified simulation is due to: (1) modest validity of the composite ($r = .35$, based both on the simulation and analytic analysis) and (2) the assumed quality of the candidate group. Any composite that gives considerable weight to g is expected to result in large values for the B-W d and severe adverse impact (Sackett & Ellingson, 1997).

The results of these two simulations are strong indications that our selection tests for police officer are inadequate to the task.

An Administrative Job Aid for Municipal Hiring Officials

Police managers often are shocked when a new selection system for entry-level police officer results in severe adverse impact. Testing consultants often (typically?) recommend using M/C tests of g for selecting police officers. When asked by municipal officials to design a selection battery to help the jurisdiction hire a diverse group of new police officers, consultants often respond by including a test with low B-W d (e.g., personality) as a weighted component in calculating a composite test score. However, this composite approach often results in severe adverse impact (as we saw in the simulation above). By the time the municipal officials learn of the severe adverse impact of their new exam, the exam has been given and graded and it is too late to make changes to the exam. To the surprise and dismay of the municipal managers, the new police officer selection system becomes a political and, perhaps, a legal liability.

A solution I propose involves giving municipal officials more information about and a larger role in designing any new employee selection system for police officers. Municipal officials can help design the entry-level testing program if given numerical estimates of the

predicted diversity in hiring and the expected level of job performance for various testing approaches. To facilitate or enable this, the municipal officials need information that is not readily available to them. I suggest testing experts provide this information in a structured fashion, as follows. In the RFP process, require the consultants to give the municipal officials numeric predictions of the expected number of diversity hires that will result from the proposed selection test and the level of job performance expected from those hired. The municipal officials can, in turn, use the information provided to decide which of various testing proposals to accept. That way, the municipal officials will be in a good position to make decisions concerning the relative managerial, legal, and political value of:

- diversity in newly hired police officers
- the cost and time required for developing and conducting the selection process
- the expected level of job performance
- practicality and transparency of the selection system

Pros:

- weighty decisions will be made by the responsible municipal officials, and
- municipal officials are less likely to be surprised after the fact by the level of adverse impact of a new hiring process.

Cons:

- municipal managers have limited expertise in psychometrics, so both they and their consultants will have to work hard to bring them up to speed on such concepts as: utility, expected level of job performance, expected level of adverse impact, and variance in these expected values,
- municipal officials may be hard pressed to provide some of the information that consultants need to make predictions of the number of diversity hires, and
- consultants will feel uncomfortable making predictions about the number of diversity hires.

A Proposal Evaluation Form, a job aid to facilitate the collection and evaluation of this information, is provided in Figure 3. (This figure may be difficult to read, so a larger version is available on the web at <http://jpwphd.com/acn>.)

I look forward to writing one or more ACN articles on other aspects of police officer hiring. I do not have all the solutions, but I do have some additional observations and ideas for ways to address the disturbing outcomes mentioned at the beginning of this article. Some of my other writings on this topic are available at <http://jpwphd.com/acn>.

Questions and comments on this column are welcome. Write me at: jpw@jpwphd.com.

Figure 3. Form Proposed to Collect Psychometric Information from Potential Consultants

Police Officer Selection System Proposal Evaluation Form				
Topic	Selection System Approach 1	Selection System Approach 2	Selection System Approach 3	Selection System Approach 4
1. Projected Number of Hires				
Projected number of whites hired				
Projected number of blacks hired				
Projected number of Hispanics hired				
Projected number of other minorities hired				
Projected number of men hired				
Projected number of women hired				
2. Projected Adverse Impact				
Projected Adverse Impact: Blacks				
Projected Adverse Impact: Hispanic				
Projected Adverse Impact: Other minorities				
Projected Adverse Impact: Women				
3. Projected Average Job Performance (fill in one of the three lines below)				
Average job performance using SAT scale (where average of all applicants is 500, standard deviation is 100)				
% of hire who will be successful on the job				
Other measure of projected job performance (described in narrative of proposal)				
4. What is the basis for ranking candidates? (narrative)				
5. Cost				
Total Cost for projected number of candidates				
Adjustment +/- for more or fewer candidates				
6. Time to implement full selection system				
7. Evaluation of Proposer (acceptable or not, plus narrative)				
Past experience				
Expertise				
References				
8. Quality of Proposal (acceptable or not, plus narrative)				
Understandable, logical, and complete (do I trust projections)				
Selection system practical and reasonable				
9. OVERALL EVALUATION (Rate on a 1-9 scale or rank from low to high)				

For more information on using this form contact:
 Dr. Joel P. Wiesen, Director
 Applied Personnel Research
 62 Candlewood Road, Scarsdale, NY 10583
projections@appliedpersonnelresearch.com
 (617) 244-8859

Instructions to Chief
 Summarize 1-5 from proposals
 Rate 6-8 numerically or with narrative
 Rate 9, considering 1-8.