

Multiple Ways to Increase Both Diversity and Expected Job Performance in Hiring and Promoting Police Officers

Joel P. Wiesen, Ph.D.

jw@jpwphd.com
(617) 244-8859

GLEAN Virtual Learning Event 10/26/2022

1

Print and Audio Links

- PowerPoints (yet to be posted)
- Audio recording (yet to be posted)
- **<http://jpwphd.com/glean2022>**

GLEAN Virtual Learning Event 10/26/2022

2

Topics of This Presentation

- Effectiveness of entry and promotional tests
- Are administrative decisions professional?
- Approaches to reduce AI that do not work
- Tests with lower validity can result in higher expected job performance
- Some real-world examples

Notation: AI = adverse impact

GLEAN Virtual Learning Event 10/26/2022

3

Learning Objective 1

- Describe two research based approaches to both improve diversity in hiring and improve expected job performance of entry-level police officers.

GLEAN Virtual Learning Event 10/26/2022

4

Learning Objective 2

- Describe two research based reasons (i.e., findings) for using tests of g on a pass-fail basis in hiring entry-level police officers.

Notation: g = cognitive ability or general mental ability (GMA)

Learning Objective 3

- Describe two major threats to the content validity of police promotional exams based on a reading list.

Learning Objective 4

- Describe two possible approaches to improving the utility of police promotional exams.

Good News and Bad News

- Look at some bad news first

Effectiveness of Hiring/Promoting

- Mathematical evaluation of the rate of job success for entry-level hires is possible based on criterion validity studies #
- Cannot evaluate success rate for promotions
 - Not one criterion-related validation study
 - At least I could not find any criterion studies
 - Any claim of success is a leap of faith
 - Content validity is based on judgments

GLEAN Virtual Learning Event 10/26/2022

9

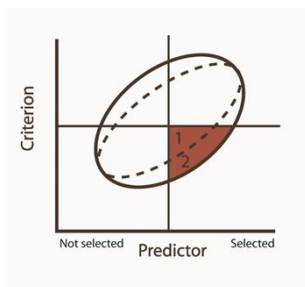
Two Types of Hiring Errors

- False Positive:
A candidate is **hired but fails on the job**
- False Negative:
A candidate could do the job but is not hired

GLEAN Virtual Learning Event 10/26/2022

10

False Positives: Two Levels of r



GLEAN Virtual Learning Event 10/26/2022

11

Low Validity of g for PO

- Meta-analysis
 - $r = .27$, over corrected
 - Corrected for predictor unreliability
 - $r = .24$, corrected
 - Corrected for all but predictor unreliability
 - Less valid with college education requirement
- Source: Aamodt (2004)
- Such exams are still in use today, widely

GLEAN Virtual Learning Event 10/26/2022

12

Two Views of $r = 0.24$

- M/C test is measuring all KSAPs that can be measured.
 - There are many false positives
- M/C test misses some needed KSAPs
 - Even more false positives due to these KSAPs

Notation: r = validity (i.e., the mathematical correlation between selection test score and job performance measure)

Many Entry Hires Will Fail on Job

- About **40% of new Police Officers** are mathematically **expected to fail on the job**
 - 60% fail the job if the test omits crucial KSAPs
- Due to low validity of entrance tests ($r = .24$) and diverse abilities among applicants
- Based on criterion studies (no leap of faith)
(Source: Wiesen, 2018)

Some Data on Promotions

- The mean score on promotional exams is often (typically?) in the 60's with S.D. < 10
- Highest scores often in the 80's
 - Missing 10+% of crucial knowledges
- Chiefs have told me that they cannot promote their best officers because they score too low on the civil service exams
- It seems the promotional system is **flawed**

Possible Solutions

- Entry-level#
 - Change weighting of test subjects or components
 - Change components
 - Use administrative decisions
- Promotion #
 - Improve exams
 - Change components
 - Train for promotion (would be a major change)

Entry-Level Hiring Approaches

- PDs have two goals
- Improve expected job performance
- Improve diversity of hires

GLEAN Virtual Learning Event 10/26/2022

17

Six Promising Approaches

- 1 Use cognitive ability tests pass/fail, if at all
- 2 Select tests based on utility, not validity
- 3 Test KSAPs w small, zero, or reverse *d*
- 4 Have testing consultants project the number of diversity hires
- 5 Recruit quality over quantity of candidates
- 6 Use hiring preferences (e.g., residency)

GLEAN Virtual Learning Event 10/26/2022

18

1 Use Tests of *g* Pass/Fail

- Psychometrics of p/f use of *g* #
- Validity of *g* over time on the job
- How much *g* is needed?
- Fairness of tests of *g*
- Real life examples of pass/fail #

GLEAN Virtual Learning Event 10/26/2022

19

Psychometrics of P/F Use of *g*

- Traditional wisdom: *g* is best
 - Perhaps all validity is due to *g*
- Traditional wisdom is changing: *g* helps #

GLEAN Virtual Learning Event 10/26/2022

20

Prevalent Wisdom Is Changing

- g is not the highest validity #
- M/C tests of g are deficient
 - Narrow measures of some facets of intelligence
- There are valid KSAPs beyond g #
- Many **false positive** hires with a test of g #
- Validity sums, not averages (usually)
- Composites usually have lower d than g

GLEAN Virtual Learning Event 10/26/2022

21

Assessment Centers vs g

- Meta-analysis of studies of **both** AC and g
- “In contrast to Schmidt and Hunter’s [1998]... reporting... **.51 for ability and .37 for ACs**, we found [2017] ... mean **validity of .22 for ability and .44 for ACs.**”
- Assessment exercises have higher validity

(Source: Sackett, Shewach, & Keiser, 2017)

GLEAN Virtual Learning Event 10/26/2022

22

Historical Validity of Interviews

- 1981: Invalid (Tenopyr)
- 1982: low validity, low reliability, & susceptible to bias (Avery & Campion)
- 1989: At least modest validity (Harris)
- 1998: Highest validity (Schmidt & Hunter)
- 2021: Highest validity (Sackett et al., 2021)

GLEAN Virtual Learning Event 10/26/2022

23

Why Interview Validity Changed?

- Shift to structured interviews
- Added range of content beyond g

GLEAN Virtual Learning Event 10/26/2022

24

g Less Valid Over Time

- GPA used as proxy for g to predict job perf.
- $r = .49$ at 1 year post college graduation
- $r = .33$ at 2-5 years post graduation
- $r = .12$ at 6 years post graduation (n.s.)
- All **corrected** r 's with job performance
(Source: Roth, BeVier, Switzer & Schippmann, 1996)

GLEAN Virtual Learning Event 10/26/2022

25

Use of g to Rank Yields High d

- d is a measure of adverse impact (AI)
- Even low weight for g causes composite AI
 - Sackett & Ellingson (1997, Table 2)
- Effective weight for AI can be different than the nominal weight in a composite
- Conclusion: use tests of g on pass-fail basis
 - **Don't let low validity, high d test drive AI**

GLEAN Virtual Learning Event 10/26/2022

26

Do Police Officers Need High g ?

- Mean IQ for police officers = 104
 - Aamodt (2004b, based on total of 4,061 POs)
 - Median and mode surely lower than mean
- Similar normative data for the Wonderlic

GLEAN Virtual Learning Event 10/26/2022

27

1994 Nassau County Exam

- Used M/C test of *reading* on pass-fail basis
- No (other) M/C test of g
- Developed by team of IO superstars
- Critiqued by Gottfredson and many others
 - “gerrymandering” “worthless” “disaster”
- **No follow up research. Why?**
 - The hires are near retirement age

GLEAN Virtual Learning Event 10/26/2022

28

Psychometric Evaluation

- g weakly predicts PO job performance
- g drives adverse impact (AI)
- Even low weight for g causes composite AI
- Other predictors have good r & smaller d
- Questions about fairness of tests of g
 - Indications of bias in job performance criteria

A Challenge to Fairness

- Conundrum: Black-White group differences
 - $d = 0.5$ for job performance
 - $d = 1.0$ for test performance

Notation: d = size of difference between means, usually expressed in z-score units

M-W Test $d = 1.0$, Job $d = 0.5$

- The Minority-White (M-W) mean score difference (d) on test score is twice the d on job performance
- This is unfair on its face, but we define fairness in a way that gets around it

Can We Explain Conundrum?

- $Y = ax + b$
- Job performance = validity (x) + constant
- $0.5 = r * 1.0$
- This seems to work when $r = 0.5$, but:
- If r not equal 0.5, formula does not work
- If r were .9, the formula would not work

Where P/F Approach Was Used

- Miami, FL
 - (E. Kraus, personal communication, 4/4/2018)
- Bridgeport, CT
 - M/C pass/fail, ranks based on oral board (Only In Bridgeport, 2015))
- Columbus, OH Police (Columbus CSC, 2014)
 - M/C and writing sample pass/fail
 - Ranks on oral/video test of problem solving

GLEAN Virtual Learning Event 10/26/2022

33

2 Select Tests Based on Utility

- Utility is the bottom line #
 - Validity is only one of 3 drivers of utility

GLEAN Virtual Learning Event 10/26/2022

34

What Is Utility

- “Projected productivity gains ... due to use of the selection procedure”
 - (SIOP *Principles*, 2018, page 33, col 1, par 4)
- We will use change in % false positives
- We will ignore cost of recruitment, testing, training, etc., and focus on job performance (for the sake of this presentation)

GLEAN Virtual Learning Event 10/26/2022

35

Utility \neq Validity

- Utility: “projected productivity gains ... due to use of the selection procedure” (SIOP, 2018)
- Validity: “evidence and theory support ... proposed uses of ... selection procedure”
 - Test scores are related to job performance
 - Validity is not a measure of job performance
- **Most valid \neq most job performance gain**

GLEAN Virtual Learning Event 10/26/2022

36

Three Variables Drive Utility

- Quality of applicants (**Q**)
 - Proportion of applicants who can do the job
 - Best way to improve expected job performance
- Number of openings and applicants
 - Selection ratio (SR)
- Validity (r , i.e. the correlation of test and job performance)
 - Cascio & Aguinis (2011, pg 328)
 - Taylor & Russell (1939)

GLEAN Virtual Learning Event 10/26/2022

37

Is Utility or Validity Primary?

- Utility and validity are different but related
- Profession seems to largely ignore utility
- **A less valid test can have higher utility**
- Selecting tests on utility may favor diversity
- Management is interested in **utility**

GLEAN Virtual Learning Event 10/26/2022

38

Now We Largely Ignore Utility

- No review of utility in test technical manuals
- Past claims of high utility poorly received
- 1970 *EEOC Guidelines* called for high utility (Guion, 2011, page 128)
- Superseded by 1978 *Uniform Guidelines*
- **But utility is the reason we test**
 - Validity is important as it contributes to utility

GLEAN Virtual Learning Event 10/26/2022

39

What Happens with Higher Q?

- We hire better people
- Less room for improvement over chance
 - Cannot do much better than hiring randomly
 - **Improvement in utility is low with high Q**

Notation: Q = % of all applicants who can do the job

GLEAN Virtual Learning Event 10/26/2022

40

Which More Important: Q, SR, r?

- In theory: all 3 are very important
- In practice: Q and SR are more easily changed than r.
 - r is hard to change
- Better SR comes with worse adverse impact
- **Takeaway: Pay attention to recruitment**
 - Need a high proportion of good applicants

GLEAN Virtual Learning Event 10/26/2022

41

Practical Implications of Q

- Can only select from among applicants
- If no good applicants, cannot hire superstars
- If all applicants great, all hires will be great
 - Random hiring will yield superstars
- **NOTE:** The above do not depend on r
- Must pay attention to quality recruitment
- Cannot recruit more after we see test scores

GLEAN Virtual Learning Event 10/26/2022

42

High Validity with Zero Utility

- A valid test can have no practical import
- If not enough candidates and all are hired
- If all candidates are outstanding
- So need to look carefully at utility
- A less valid test can have higher utility #

GLEAN Virtual Learning Event 10/26/2022

43

Less Valid Test w/ Higher Utility

Test	r	Q	Max. Possible Change in Utility (1.0-Q)
g	0.24	0.95	0.05
Personality	0.15	0.5	0.50

GLEAN Virtual Learning Event 10/26/2022

44

Formulas

- Formulas available in Wiesen (2018)
 - Standardized mean score difference
 - Pooled estimated population standard deviation

3 Test Valid Abilities w/ Low d

- Must rank on some valid test
- Preferably a test with low d #

APA: Fairness Overrides Validity

- “If ... excluding some components ... has a noticeable impact on selection rates for groups ... the intended interpretation of test scores ... would be **rendered invalid**.”
 - Joint *Standards* (AERA, APA, NCME, 2014, page 21, col 1, par 1, emphasis added)
- Joint *Standards*: not measuring low d
KSAP invalidates any validity

Low d , Job-Related Abilities

- Face recognition and memory #
- Creative problem solving
- Oral communication
- Conscientiousness, integrity, etc. #
- Threat detection
- Depth perception
- Structured oral exams (tapping non g areas)#

Face Recognition/Memory

- Face recognition: compare different views
- Face memory: recall, perhaps w recognition
- Job related
 - Recognize lost persons and suspects
- Content, use faces that mirror:
 - the community
 - offenders

GLEAN Virtual Learning Event 10/26/2022

49

Face Recognition: d

- Expect **reverse** impact
- Remembering and identifying minority faces is easier for members of that minority group
 - e.g., Levin (2000)

GLEAN Virtual Learning Event 10/26/2022

50

Personality More Valid with Time

- $r = .18$ to $r = .45$ for GPA, from year 1 to year 7 of medical school

(Lievens, Ones & Dilchert, 2009)

GLEAN Virtual Learning Event 10/26/2022

51

Where Low d KSAPs Used

- Many jurisdictions/consultants measure non-g areas
- Some jurisdictions use oral exams to rank
 - Miami, FL
 - Bridgeport, CT
 - Columbus, OH
- **No** PDs use face recognition/memory
 - As far as I know

GLEAN Virtual Learning Event 10/26/2022

52

Unrealistic Claims of Fairness

- Issue: We claim M/C tests of g are fair despite evidence that job criteria are biased
 - Women paid less than men for same work
 - Short people paid less than tall
 - Homely people paid less than handsome
- Our tests fairly predict biased criteria
 - Perhaps this shows our tests are biased
- d s for test and job performance are logical #

GLEAN Virtual Learning Event 10/26/2022

53

4 Project Number of Diversity Hires

- Empower the PD to choose testing approach
- Empower managers to make tradeoffs #

GLEAN Virtual Learning Event 10/26/2022

54

Watchword “No Surprises”

- Managers do not like surprises
- Projections attempt to avoid surprises
- Increase role of PD in making decisions
- Improve PD decisions on consultant hiring

GLEAN Virtual Learning Event 10/26/2022

55

Focus the Proposal Comparison

- Help the PD evaluate proposed selection systems
- Increase the role of PD in deciding about tradeoffs between conflicting goals #
 - Cost and features
 - Validity
 - Utility
 - Diversity of hires, evaluated numerically

GLEAN Virtual Learning Event 10/26/2022

56

Focus the Proposal Comparison

- Focus decision-making on likely results #
- Hiring projections essential for managers
- Managers are responsible for the testing decisions, so they should make the decisions

Require Numeric Projections

- Project number of hires by protected class
- Project adverse impact
- Project utility
- Describe cost and timeline
- Need to clarify all the above #

Proposal Comparison Form

Police Officer Selection System Proposal Evaluation Form		
Topic	Selection System Approach 1	Selection System Approach 2
1. Projected Number of Hires		
Projected number of whites hired		
Projected number of blacks hired		
Projected number... etc.		

Proposal Comparison Form

Police Officer Selection System Proposal Evaluation Form		
Topic	Selection System Approach 1	Selection System Approach 2
2. Projected Adverse Impact		
Ad Impact: black		
Ad Impact: Hispanic		
Etc.		

Proposal Comparison Form

Police Officer Selection System Proposal Evaluation Form		
Topic	Selection System	
	Approach 1	Approach 2
3. Projected Job Performance		
Option 1. Mean using SAT-type scale		
Opt. 2. % hires who will be successful		
Opt. 3. Other job performance measure		

Proposal Comparison Form

- Full form available on my website
– Wiesen (2018)

5 Recruit Quality Candidates

- Recruitment and validity have about equal impact on utility in usual situations
- In general, test validity has not improved in the past several decades (Cascio & Aguinis, 2008, page 141)
- Easier to improve recruitment than validity

6 Residency Preference; HS GPA

- Points for residents of the municipality #
– Knowledge of the community
- Points for working in the municipality
- Credit for high school GPA
– HS GPA is as valid as a test of g #
- Rank based on high school GPA

Grant Residency Preference

- Residency preference is historic
 - Absolute
 - Points added to final exam score
- Absent residency preference, majority of force can come from outside the city
- Use where poor city schools are surrounded by rich suburban schools
 - Unequal educational systems

GLEAN Virtual Learning Event 10/26/2022

65

Use High School Rank

- But schools may have different standards
- However, a study showed $r = .91$ between GPA standardized within school or across schools and job performance
 - Sternberg (2006, page 331, col. 1, par. 3) #

GLEAN Virtual Learning Event 10/26/2022

66

Use of High School Rank

- High school GPA and SAT [g?] scores both predict college GPA well and about equally
 - Schmitt et al. (2007, Table 4)
 - Berry & Sackett (2009, Table 1)
 - Sternberg (2006, Table 7)
- Education attainment has lower levels of adverse impact than tests of GMA
 - Berry, Gruys & Sackett (2006)

GLEAN Virtual Learning Event 10/26/2022

67

Administrative Decisions Are Professionally Acceptable

- Joint *Standards* do **not** require the most valid selection procedure.
 - “Where ... validity exists, the decision as to whether to actually administer a particular test generally takes additional considerations into account. These include ... weighing of any negative consequences against the positive consequences of test use.” (AERA et al., 2014, pg 11, col 1, par 2; also see pg 21, col 2, par 4).

GLEAN Virtual Learning Event 10/26/2022

68

Administrative Decisions Are Professionally Acceptable

- SIOP *Principles* do **not** require the most valid selection procedure.
 - “... decisions ... driven by the goals of the organization and may be based on factors such as ... diversity goals of the organization ...”
 - “The goals of the organization may favor a particular alternative.”

GLEAN Virtual Learning Event 10/26/2022

69

Administrative Decisions Are Professionally Acceptable

- “For example, some organizations ... use a cutoff score rather than rank ordering to increase workforce diversity, recognizing that a reduction also may occur in job performance and utility.”
- “Whatever the decision, the testing professional should document the rationale for it.”
- *Principles*, page 32, col 2

GLEAN Virtual Learning Event 10/26/2022

70

Non-Testing Approaches

- Our expectations for tests are too high
 - Our tests are not all that good
 - We need to go beyond testing to improve PDs
- Non-testing approaches to improve job performance
 - Training #

GLEAN Virtual Learning Event 10/26/2022

71

Time to Train POs

- USA: Average of 21 **weeks** in academy
- Europe: Several **years** to become a PO
- Average # hours required to be a **barber** is greater than # hours required to be a PO
- **Nurses train for 2-4 years**. Why not POs?

Source: <https://www.cbsnews.com/news/police-training-weeks-united-states/>

Source: <https://www.cnn.com/2016/09/28/us/jobs-training-police-trnd/index.html>

GLEAN Virtual Learning Event 10/26/2022

72

Approaches to Promotion

- No criterion-related validation studies for police promotion
 - This is a serious limitation
 - How can we tell if our exams are working?
 - If you know of any such studies, please tell me

GLEAN Virtual Learning Event 10/26/2022

73

Improve Promotional Exams

- Definition of a good item #
- Item protests subvert item quality #
- Knowledge of law items #
- Definition items #

GLEAN Virtual Learning Event 10/26/2022

74

Two Definitions of a Good Item

- No protests or challenges #
 - Quotes straight from textbook
- Tests ability do the promotional job
 - Extrapolate from textbook
 - Apply knowledge to new situations

GLEAN Virtual Learning Event 10/26/2022

75

Item Protests Subvert Item Quality

- Lay body evaluates item protests
- Easiest way to defend an item is to show it is taken directly from a source document
- Reading lists can subvert item quality if:
 - Result in items that quote from sources #
 - Test recognition of wording, not understanding
- Use reading lists with care

GLEAN Virtual Learning Event 10/26/2022

76

Item Protests Subvert Item Quality

- Items with verbatim quotes from sources measure recognition of wording
- Such items don't measure application of K
- Better: Use more job simulation questions
 - Rely on SMEs to extrapolate from textbook

GLEAN Virtual Learning Event 10/26/2022

77

Knowledge of Law Items

- Law items often are exact replication of a case
- No deviation from the court case because no one knows what a court may rule if the facts were somewhat different
- But this omits exactly what a PO or Sgt needs to do to perform the job: apply the law to new situations

GLEAN Virtual Learning Event 10/26/2022

78

Minimize Use of Definition Items

- Definition items are easy to write
- K of definition is only weakly related to application of knowledge
- Avoid definition items, in general
- Use more job simulation questions

GLEAN Virtual Learning Event 10/26/2022

79

Assumptions Revisited

- Compensatory grading is illogical #
- 100 items is long enough #
- Are claims of fairness realistic? #

GLEAN Virtual Learning Event 10/26/2022

80

Compensatory Grading Illogical

- Grade is based on # correct
- Tests cover many unrelated topics
- Can promote person with gaps in KSAPs
 - Strong in law and weak in criminal investigation
- Solution: Consider multiple passing points for crucial KSAPs

GLEAN Virtual Learning Event 10/26/2022

81

Our Tests Are Too Short

- Test outline topics with only 1 or 2 items
- Few items \Rightarrow unreliable measure
- Unreliable measure \Rightarrow invalidity
- Solution: longer tests

GLEAN Virtual Learning Event 10/26/2022

82

MC Tests Do Not Test Creativity

- Creative problem solving is important
- Applying knowledge to new situations imp.
- M/C tests test recognition of solution
- M/C does not test for thinking of a solution
- Solution: More test modes

GLEAN Virtual Learning Event 10/26/2022

83

Promotion to Sergeant & Above

- Problems:
 - Low mean M/C test scores for all candidates #
 - M/C tests often have adverse impact
 - Emphasis on rote memorization
- Solutions:
 - Assure knowledge in other ways
 - Do not rank based on M/C tests
 - Rank on valid KSAPs other than a M/C test

GLEAN Virtual Learning Event 10/26/2022

84

Low Mean Scores on M/C tests

- What do the M/C tests measure?
 - Important police knowledges
 - Must know, cannot look up on the job
 - Thousands of pages of SOPs, law, etc.
- Mean scores often low (in the 60's)
- We promote people with critical K gaps
- Incumbents often fail a promotion test for their job

GLEAN Virtual Learning Event 10/26/2022

85

Promoted w Low Exam Scores

- Often, highest test score is in 80's
- Items chosen to be important, even crucial
 - Miss 10+ crucial items
- Possible explanations
 - No training for new job (esp. promotions)
 - Exams not related to (most) job duties
- Implication: high false positive rate

GLEAN Virtual Learning Event 10/26/2022

86

Incumbents Often Fail a Test for Their Own Job

- In one jurisdiction, candidates for sergeant and lieutenant take the same 80 item test
 - Lt. applicants take an additional 20 items
- Applicants for Lt have studied
- Applicants for Lt have 3+ years on the job
- 12% of Sergeants fail the Sgt exam
- 15% of minority Sgts fail the Sgt exam

GLEAN Virtual Learning Event 10/26/2022

87

Why Do Incumbents Fail?

- All incumbent Sgts scored high enough to be promoted in the past
- Did they forget what they once knew?
- Did years on the job cause them to forget?
- Low reliability tests? #
- Promoted due to guessing (chance)
- No good reason for incumbent failure rates

GLEAN Virtual Learning Event 10/26/2022

88

Rater Reliability Inflated

- Typically, we measure within panel rater reliability and it is high, after discussion
- Between panel reliability is much lower
- High reliability within panels (.8's, .9's)
- Low reliability between panels (.4's to .7's)
- Solutions: Better rater training; types of exercises that led to low and high reliability

Re-Envision Promotional Tests

- Now: No training for promotion
- Now: We expect Police Officers to teach themselves how to be Sergeants

Re-Envision Promotional Tests

- Now: Issue reading list and then test
 - Rules, procedures, and guidelines, and sometimes police science or other textbooks
- Books teach principles but give no practice
 - Interpersonal aspects of supervision
 - Problem solving
 - Strategy & tactics
 - Fast size-up and decision making

Re-Envision Promotional Tests

- Future: Train to criteria on written sources
 - SOPs
 - Directives
 - Law
 - Textbooks (practical in orientation)
- Train to criteria on strategy and tactics #
- Train to criteria on supervision, leadership

Ideas for Implementation

- PD developed courses in strategy/tactics
- Self-paced learning of sources
- Mastery tests to earn course credit
- Retake course if needed
- Many short courses on the sources

GLEAN Virtual Learning Event 10/26/2022

93

More Use of Job Performance

- Commonly accepted that past job performance is best predictor of future job performance
- Typical vs. maximal level of performance

GLEAN Virtual Learning Event 10/26/2022

94

Better Measures of Job Performance

- Body worn cameras allow supervisors to observe job performance as never before
- Unfortunately, PDs now only review body cams if there is a serious incident
- Missing the opportunity to evaluate everyday interactions with the community

GLEAN Virtual Learning Event 10/26/2022

95

If We Are Too Conservative

- If we continue as we have in the past:
- The adverse impact will continue unabated
- The poor PO job performance will continue

GLEAN Virtual Learning Event 10/26/2022

96

The Field Suffers from Secrecy

- Consultants refine their products
- Best work is not shared
 - BARS
 - M/C items
 - Work sample items
- Field advances slowly without sharing
- IPAC goal is to share work

GLEAN Virtual Learning Event 10/26/2022

97

Learning Objective 1 w/ Answers

- ***Describe two research based approaches to both improve diversity in hiring and improve expected job performance of entry-level police officers.***
- Use M/C test of *g* P/F if at all
- Measure KSAPs with low or reverse impact (e.g., face recognition, creativity, conscientiousness)

GLEAN Virtual Learning Event 10/26/2022

98

Learning Objective 2 w/ Answers

- ***Describe two research based reasons (i.e., findings) for using tests of *g* on a pass-fail basis in hiring entry-level police officers.***
- Validity of *g* decreases with time (per some studies)
- Including *g* in a composite drives AI
- Validity of personality increases with time
 - Don't let a less valid measure drive AI

GLEAN Virtual Learning Event 10/26/2022

99

Learning Objective 3 w/ Answers

- ***Describe two major threats to the content validity of police promotional exams based on a reading list.***
- Questions tend to measure memory of facts rather than application of knowledge
- Few test questions for some knowledges

GLEAN Virtual Learning Event 10/26/2022

100

Learning Objective 4 w/ Answers

- ***Describe two possible approaches to improving the utility of promotional exams.***
- Measure (carefully) more KSAPs (e.g., strategy and tactics, problem solving)
- Use a cut point for each KSAP

Takeaways

- Recruitment can improve utility more than testing
- A low validity test can have high utility
- A high validity test can have low utility
- *g* is not the best predictor of job performance

Topics Not Covered

- Validity sums (validity does not average)
- Adding a low validity test improves validity
- Numeric examples
- Other ways to reduce adverse impact
- Other real life applications
- Some of this is on my website:
<http://jpwphd.com/glean2022>

Question and Answer Time

- Related:
– <http://jpwphd.com/papers>
- Contact me after the conference, too.
Joel P. Wiesen, Ph.D,
jpw@jpwphd.com
(617) 244-8859 (land line)
- **Reference section is on the website**

References

- Aamodt, M. G. (2004a). *Research in Law Enforcement Selection*. Boca Raton, FL: Brown Walker Press.
- Aamodt, M. G. (2004b). *Law Enforcement Selection: Research Summaries*. Washington, D.C.: Police Executive Research Forum.

References

- AERA, APA, NCME. (2014) *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

References

- Arvey, R. D. & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology*, 35, 281-322.

References

- Berry, C. M., Gruys, M. L. & Sackett, P. R. (2006). Educational Attainment as a Proxy for Cognitive Ability in Selection: Effects on Levels of Cognitive Ability and Adverse Impact. *Journal of Applied Psychology*, 91, 696-705.

References

- Berry, C. M. & Sackett, P. R. (2009) Individual differences in course choice result in underestimation of the validity of college admissions systems. *Psychological Science, 20*, 822-830.

References

- Cascio, W. F. & Aguinis, H. (2008). Staffing Twenty-first-century Organizations. *The Academy of Management Annals, 2*, 133–165.
- Cascio, W. F. & Aguinis, H. (2011). *Applied Psychology in Human Resource Management*. Boston: Pearson.

References

- Columbus Civil Service Commission (2014). *Study Guide for the City of Columbus, Ohio Police Officer Entry Level Examination*. Downloaded 4/5/2017 from <https://www.columbus.gov/civilservice/uniformed-police-series/Police-Officer/>

References

- Guion, R. M. (2011). *Assessment, Measurement, And Prediction For Personnel Decisions* (2nd ed.) Bowling Green State University. New York: Routledge.

References

- Levin, D. T. (2000) Race as a Visual Feature: Using Visual Search and Perceptual Discrimination Tasks to Understand Face Categories and the Cross-Race Recognition Deficit. *Journal of Experimental Psychology: General*, 129, 559-574.

GLEAN Virtual Learning Event 10/26/2022

113

References

- Harris, M. M. (1989) Reconsidering the employment interview: a review of recent literature and suggestions for future research. *Personnel Psychology*, 42, 691-726.

GLEAN Virtual Learning Event 10/26/2022

114

References

- Lievens, F., Ones, D. S. & Dilchert, S. (2009). Personality Scale Validities Increase Throughout Medical School. *Journal of Applied Psychology*, 94, 1514–1535.

GLEAN Virtual Learning Event 10/26/2022

115

References

- Only In Bridgeport (Web Magazine). (2015, July 27). *Finch Announces New List For Police Recruits*. Retrieved from <http://onlyinbridgeport.com/wordpress/finch-announces-new-list-for-police-recruits/>

GLEAN Virtual Learning Event 10/26/2022

116

References

- Roth, P. L., BeVier, C. A., Switzer III, F. S. & Schippmann, J. S. (1996). Meta-Analyzing the Relationship Between Grades and Job Performance. *Journal of Applied Psychology, 81*, 548-556.

References

- Sackett, P. R. & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50*, 707-721.

References

- Sackett, P. R., Shewach, O. R. & Keiser, H. N. (2017). Assessment centers versus cognitive ability tests: challenging the conventional wisdom on criterion-related validity. *Journal of Applied Psychology, 102*, 1435–1447.

References

- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2021, December 30). Revisiting Meta-Analytic Estimates of Validity in Personnel Selection: Addressing Systematic Overcorrection for Restriction of Range. *Journal of Applied Psychology*. Advance online publication. <http://dx.doi.org/10.1037/apl0000994>

References

- Schmitt, N., Oswald, F. L., Kim, B. H., Imus, A., Merritt, S., Friede, A. & Shivpuri, S. (2007). The Use of Background and Ability Profiles to Predict College Student Outcomes. *Journal of Applied Psychology*, 92, 165–179.

References

- SIOP (2018). *Principles for the Validation and Use of Personnel Selection Procedures*, 5th ed. Bowling Green, OH: Author.

References

- Sternberg, R. G. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence*, 34, 321–350.

References

- Taylor, H. C, & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology*, 23, 565-578.

References

- Tenopyr, M. L. (1981). The realities of employment testing. *American Psychologist*, 36, 1120-1127.

References

- Wiesen, J. P. (2018). *Tools to Increase Diversity, Utility, and Validity in Hiring Police Officers*. Presented at the 33rd Annual Convention Society for Industrial and Organizational Psychology, Chicago, IL.