

# The 3 Best Psychometric Approaches to Maximize Diversity in Hiring

Joel P. Wiesen, Ph.D.

[w@jpwphd.com](mailto:w@jpwphd.com) & <http://jpwphd.com/papers>

46<sup>th</sup> Annual Conference  
Society for Police and Criminal Psychology  
Online; November 5-7, 2020

Wiesen (2020) - Society for Police and Criminal Psychology Conference

1

---

---

---

---

---

---

---

---

## Learning Objectives

- Describe two research-based approaches to both improving diversity in hiring and improving expected job performance of police officers.
- Describe two research-based reasons for using tests of *g* on a pass-fail basis.

Wiesen (2020) Society for Police and Criminal Psychology Conference

2

---

---

---

---

---

---

---

---

## Related Print and Audio Links

<http://jpwphd.com/papers>

Feel free to contact me about this topic for any reason:

- [jpw@jpwphd.com](mailto:jpw@jpwphd.com)
- (617) 244-8859 (land line)

Wiesen (2020) Society for Police and Criminal Psychology Conference

3

---

---

---

---

---

---

---

---

## Define “Best” in the Title

- Title: The 3 **Best** Psychometric Approaches to Maximize Diversity in Hiring
- Psychometric support
- Expect to reduce  $d$  (and adverse impact)
- Expect to **improve job performance**
- Practical
- Have been used operationally



Wiesen (2020) Society for Police and Criminal Psychology Conference

4

---

---

---

---

---

---

---

---

## 3 Best Psychometric Approaches

- (1) Use tests of general cognitive ability ( $g$ ) on a pass/fail basis
- (2) Select tests based on utility, not validity
- (3) Test specific job-related cognitive abilities and other characteristics that show small ethnic group differences (i.e.,  $d$  near zero)



Wiesen (2020) Society for Police and Criminal Psychology Conference

5

---

---

---

---

---

---

---

---

## Topics for Each Approach

- Need
- Psychometric support
- Effect on  $d$  and adverse impact
- Why to expect improved job performance
- Practical considerations
- Where approach was used operationally



Wiesen (2020) Society for Police and Criminal Psychology Conference

6

---

---

---

---

---

---

---

---

## Focus on $d$ , not Adverse Impact

- Standardized B-W mean score difference ( $d$ )
- Adverse impact (AI) ratio bounces around
  - Influenced by exact number hired, small Ns
- $d$  is a more stable measure than the AI ratio



Wiesen (2020) Society for Police and Criminal Psychology Conference

7

---

---

---

---

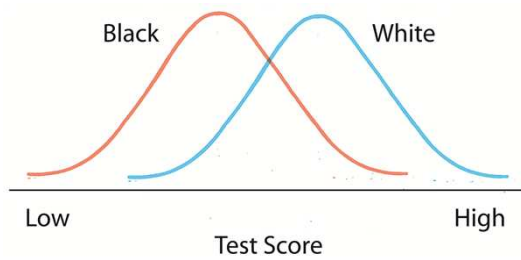
---

---

---

---

## One Standard Dev. Difference



Wiesen (2020) Society for Police and Criminal Psychology Conference

8

---

---

---

---

---

---

---

---

## (1) Use $g$ on a pass/fail basis

- Need: Tests of  $g$  have high  $d$  (about 1.0)
- $d$  of 1 causes severe AI
- Even smaller  $d$  (e.g. .3) causes AI
- $g$  in a composite causes high  $d$ , severe AI
  - $d$ 's do **not** average
    - Sackett & Ellingson (1997)



Wiesen (2020) Society for Police and Criminal Psychology Conference

9

---

---

---

---

---

---

---

---

## Project Number of Hires

- Assumptions
- Normal distribution
- 10,000 applicants
  - 9,000 White; 1,000 Minority
- $d = 1.0$
- Selection ratio = .05 (i.e., 500 openings)
- What if we double number of applicants?



---

---

---

---

---

---

---

---

## Projected Minority Hiring

Total # of Applicants	# of Minority Applicants	# of Minority Hires	Adverse Impact Ratio
10,000	1,000	4	.08
20,000	2,000	3	.06

Selection ratio = .05 (i.e., 5%);  $d=1.0$



---

---

---

---

---

---

---

---

## Psychometrics of p/f use of g

- Traditional wisdom:  $g$  is best
- Traditional wisdom is changing:  $g$  helps



---

---

---

---

---

---

---

---

## Traditional Wisdom

- $g$  has the highest validity
- There is not much beyond  $g$
- Low validity tests dilute the validity of  $g$
- Can select good employees with test of  $g$
- Strong risk of increased  $d$  with a composite



---

---

---

---

---

---

---

---

## Prevalent Wisdom Is Changing

- $g$  is not the highest validity
- Tests of  $g$  are deficient
  - Deficient measures of intelligence
  - Valid KSAPs beyond  $g$
- Validity sums, not averages (usually)
- Many **false positive** hires with a test of  $g$
- Composites usually have lower  $d$  than  $g$



---

---

---

---

---

---

---

---

## Low Validity of $g$ for PO

- Aamodt (2004a) meta-analysis
  - $r = .27$ , over corrected
    - Over corrected, for predictor unreliability
  - $r = .24$ , **corrected**
    - Corrected for all but predictor unreliability



---

---

---

---

---

---

---

---

## *g* Less Valid with Passing Years

- GPA as proxy for *g* (and more)
- $r = .49$  at 1 year post college graduation
- $r = .33$  at 2-5 years post graduation
- $r = .12$  at 6 years post graduation
- All corrected  $r$ 's
- Roth, BeVier, Switzer & Schippmann (1996)



Wiesen (2020) Society for Police and Criminal Psychology Conference

16

---

---

---

---

---

---

---

---

## Assessment Centers vs *g*

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



Wiesen (2020) - Society for Police and Criminal Psychology Conference

17

---

---

---

---

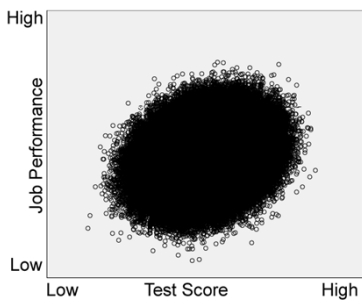
---

---

---

---

## M/C Predictive Validity, $r = .24$



Wiesen (2020) Society for Police and Criminal Psychology Conference

18

---

---

---

---

---

---

---

---

## Hiring Errors

- False Positive:  
A candidate cannot do the job but is hired.
- False Negative:  
A candidate could do the job but is not hired.



---

---

---

---

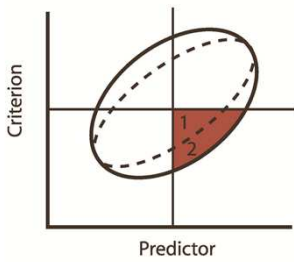
---

---

---

---

## False Positives: Two Levels of $r$



---

---

---

---

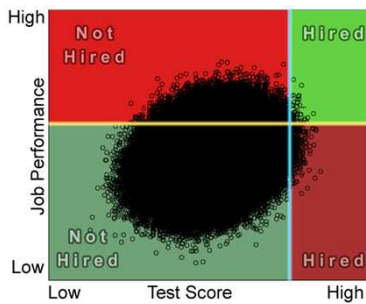
---

---

---

---

## Decisions, Right and Wrong



---

---

---

---

---

---

---

---

## Psychometric Evaluation

- Do tests of *g* result in hiring of good POs?
- What is the false positive rate?
  - We will not discuss the false negative rate
- (Analyses of expected mean job performance yield basically the same conclusions as the simpler P/F analyses.)



---

---

---

---

---

---

---

---

## False Positive Rate of *g*

- 38% false positives (62% true positives)
  - With plausible assumptions of *r* and % hired
- But 34% of the 62% are deficient on non-*g*
- So, reduce the 62% by 34% = 41%
  - $.62 \times (1-.34) = .41$
- **59% false positives** (41% true positives)
  - (c.f., Wiesen, 2018)



---

---

---

---

---

---

---

---

## Use of *g* to Rank Yields High *d*

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



---

---

---

---

---

---

---

---



## 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



---

---

---

---

---

---

---

---

## 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



---

---

---

---

---

---

---

---

## 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)
- So, the joint *Standards* say lack of fairness invalidates any indications of validity



---

---

---

---

---

---

---

---

## Better Job Performance, Lower $d$

- Allows ranking based on other valid tests
  - Not hard to create battery with  $r > .24$
  - (We will discuss utility next.)
- The test with the highest  $d$  and low validity should not drive ranking and hiring
  - Using  $g$  on a P/F basis will avoid  $g$  causing severe AI



---

---

---

---

---

---

---

---

## Where Approach Was Used

- Miami, FL  
(E. Kraus, personal communication, 4/4/2018)
- Bridgeport, CT
  - M/C pass/fail; Ranks based on oral board  
(Bridgeport Civil Service Commission, 2015)
- Columbus, OH Police (Columbus CSC, 2020)
  - M/C and writing sample pass/fail
  - Ranks on oral/video test of problem solving



---

---

---

---

---

---

---

---

## Practicality

- Easy to implement
- Easy to describe
- Candidate acceptance is good



---

---

---

---

---

---

---

---

## (2) Select Tests Based on Utility

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



---

---

---

---

---

---

---

---

## 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)



---

---

---

---

---

---

---

---

## Utility ≠ Validity

- Utility: “projected productivity gains ... due to use of the selection procedure”
- Validity: “evidence and theory support ... proposed uses of ... selection procedure”
  - Tests scores are related to job performance
  - Validity is not a measure of job performance
- **Most valid ≠ most productivity gain**



---

---

---

---

---

---

---

---

## Is Utility or Validity Primary?

- Utility and validity are not identical
- 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**



---

---

---

---

---

---

---

---

## Now, We Largely Ignore Utility

- No review of utility in test technical manual
- Past claims of high utility poorly received
- 1970 *EEOC Guidelines* called for high utility (Guion, 2011, page 128)
- Superseded by 1978 *Uniform Guidelines*
  - Business necessity not interpreted as utility
- But **utility is the reason we test**
  - Validity is important as it contributes to utility



---

---

---

---

---

---

---

---

## Three Variables Drive Utility

- Quality of applicants (**Q**)
  - Proportion of applicants who can do the job
- Number of openings and applicants
  - Selection ratio (SR)
- Validity (**r**)
  - Cascio & Aguinis (2011, pg 328)
  - Taylor & Russell (1939)



---

---

---

---

---

---

---

---

## Less Valid Test w/ Higher Utility

- Test of  $g$ :  $r = .24$ ,  $Q = .95$ 
  - i.e., 95% of applicants have the  $g$  to do the job
  - e.g., PD that requires a college degree to apply
- Test of personality :  $r = .15$ ,  $Q = .5$
- Max possible utility of  $g = 5\%$
- Max possible utility of personality = 50%



Wiesen (2020) - Society for Police and Criminal Psychology Conference

37

---

---

---

---

---

---

---

---

## Evaluate Approach 2

- Psychometric support: Just discussed
- $d$ : Lower depending on test chosen to rank
- Job performance: Shown by utility analyses
- Practical: Depends on test chosen to rank
- Where used: When use M/C test of  $g$  p/f?
- Conclusion  
Select tests based on utility, not validity



Wiesen (2020) Society for Police and Criminal Psychology Conference

38

---

---

---

---

---

---

---

---

## (3) Test Valid Abilities w/ Low $d$

- Need: Must rank on some valid test
- Preferably a test with low  $d$



Wiesen (2020) Society for Police and Criminal Psychology Conference

39

---

---

---

---

---

---

---

---

## Low *d*, Job-Related Abilities

- Face recognition and memory
- Creative problem solving
- Oral communication
- Conscientiousness, integrity, etc.
- New ways to measure intelligence
- Structured oral exams



Wiesen (2020) - Society for Police and Criminal Psychology Conference

40

---

---

---

---

---

---

---

---

## Face Recognition: Definition

- Face recognition and memory
- Recognize lost persons and perps
- Use faces that mirror the community
- Use faces that mirror the offenders



Wiesen (2020) - Society for Police and Criminal Psychology Conference

41

---

---

---

---

---

---

---

---

## Face Recognition: Validity

- No predictive validity studies (that I know)
- I expect face memory/recognition would be supported by content validity
  - Likely related to various job tasks
  - Recognizing perps
  - Recognizing other persons



Wiesen (2020) - Society for Police and Criminal Psychology Conference

42

---

---

---

---

---

---

---

---

## Face Recognition: *d*

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



Wiesen (2020) - Society for Police and Criminal Psychology Conference

43

---

---

---

---

---

---

---

---

## Creative Problem Solving

- Cannot fully measure creativity with a M/C test
  - Only open-ended questions allow for original responses



Wiesen (2020) - Society for Police and Criminal Psychology Conference

44

---

---

---

---

---

---

---

---

## Creative Problem Solving: *d*

- $r = .07$  to  $.29$  for creativity and *g*
  - e.g., Kim (2005), Sternberg (2006, Tables 9, 11.1, 11.2)
- Low or zero *d*
  - “...most studies have found that different ethnicities perform comparably on creativity tasks...”
  - (Kaufman 2010)



Wiesen (2020) - Society for Police and Criminal Psychology Conference

45

---

---

---

---

---

---

---

---

## Oral Communication

- Low  $d$  for candidates for law enforcement
  - Hausknecht, Trevor & Farr (2002)



---

---

---

---

---

---

---

---

## Integrity Tests

- $d$  around zero for race
  - Ones & Viswesvaran (1998)
- Validity high ( $r = .41$ )
  - Highest incremental validity over  $g$
  - Schmidt & Hunter (1998)



---

---

---

---

---

---

---

---

## Personality More Valid w/ Time

- $r = .18$  to  $r = .45$ , year 1 to year 7 of med school
  - Lievens, Ones & Dilchert (2009).



---

---

---

---

---

---

---

---



## New Ways to Test Intelligence

- There are some newer ways to test intelligence that show lower  $d$  values.
  - e.g., Agnello, Ryan, Yusko (2015)
- There are facets to  $g$ , some with smaller  $ds$
- Facets of  $g$  are not equally valid for various different jobs
  - e.g., Wee, Newman & Joseph (2014)



Wiesen (2020) Society for Police and Criminal Psychology Conference

49

---

---

---

---

---

---

---

---

## Structured Oral Exam, In General

- Highest validity of all tests,  $r = .57$ 
  - Aamodt (2016, Table 5.2, page 194)
- $d$  of zero
  - Levashina, Hartwell, Morgeson & Campion (2014, Table 3, page 254)



Wiesen (2020) Society for Police and Criminal Psychology Conference

50

---

---

---

---

---

---

---

---

## Evaluate Job Performance

- Cover more abilities required by the job
- So, expect improved job performance



Wiesen (2020) - Society for Police and Criminal Psychology Conference

51

---

---

---

---

---

---

---

---

## Evaluate Practicality

- Job related – must rely on content validity
- No meta-analyses for police for non-*g* tests
- There are existing tests for these abilities



Wiesen (2020) - Society for Police and Criminal Psychology Conference

52

---

---

---

---

---

---

---

---

## Where 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



Wiesen (2020) - Society for Police and Criminal Psychology Conference

53

---

---

---

---

---

---

---

---

## Summary

- Don't let low *r*, high *d* tests drive ranking
- New findings negate old findings
  - *g* **not** most valid, per Sackett et al, 2017
  - *g* validity shrinks with time
  - Personality validity grows with time
- Without new approaches, few minority hires



Wiesen (2020) Society for Police and Criminal Psychology Conference

54

---

---

---

---

---

---

---

---

## If We Are Too Conservative

- If we continue as we have in the past, the adverse impact will continue unabated.



Wiesen (2019) IPAC Conference

55

---

---

---

---

---

---

---

---

## Learning Objective 1

Describe two research based approaches to both improving diversity in hiring and improving expected job performance of police officers.

- Test valid abilities with reverse or low impact (e.g., face memory, personality)
- Use valid test modes with low impact (e.g., oral board, assessment center)



Wiesen (2020) Society for Police and Criminal Psychology Conference

56

---

---

---

---

---

---

---

---

## Learning Objective 2

Describe two research based reasons for using tests of  $g$  on a pass-fail basis.

- Low validity of  $g$  for police officer job
- Validity of  $g$  decreases with time
- Test abilities w more utility & smaller  $d$
- Not let highest  $d$  test drive ranking & AI



Wiesen (2020) Society for Police and Criminal Psychology Conference

57

---

---

---

---

---

---

---

---

## Q&A's

- Call or write me anytime to talk about this
- (617) 244-8859
- [jw@jpwphd.com](mailto:jw@jpwphd.com)



---

---

---

---

---

---

---

---

## 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

- Aamodt, M. G. (2016) *Industrial-Organizational Psychology: An Applied Approach* (8th ed.) Boston: Cengage Learning.



---

---

---

---

---

---

---

---

## References

- Agnello, P., Ryan, R. & Yusko, K. P. (2015) Implications of modern intelligence research for assessing intelligence in the workplace. *Human Resource Management Review* 25, 47–55.



---

---

---

---

---

---

---

---

## References

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.



---

---

---

---

---

---

---

---

## References

- Bridgeport Civil Service Commission (2015). *Announcement of Open Competitive Examination #2330 for Police Officer*.
- Cascio, W. F. & Aguinis, H. (2011) *Applied Psychology in Human Resource Management* (7th ed.) Saddle River, NJ: Prentice Hall.



---

---

---

---

---

---

---

---

## References

- Columbus Civil Service Commission (2020). *The Selection Process*. Downloaded 9/9/2020 from <https://www.columbus.gov/police-officer/selection-process/>



---

---

---

---

---

---

---

---

## References

- Guion, R. M. (2011). *Assessment, Measurement, and Prediction for Personnel Decisions (2<sup>nd</sup> ed)*. New York: Routledge.



---

---

---

---

---

---

---

---

## References

- Hausknecht, J. P., Trevor, C. O. & Farr, J. L. (2002). Retaking ability tests in a selection setting: Implications for practice effects, training performance, and turnover. *Journal of Applied Psychology*, 87, 243-254.



---

---

---

---

---

---

---

---

## References

- Kaufman, J. C. (2010). Using creativity to reduce ethnic bias in college admissions. *Review of General Psychology, 14*, 189–203.
- Kim, K. H. (2005). Can only intelligent people be creative?; A meta-analysis. *The Journal of Secondary Gifted Education, 16*, 57–66.

Wiesen (2020) Society for Police and Criminal Psychology Conference

67

---

---

---

---

---

---

---

---

## References

- Levashina, J., Hartwell, C. J., Morgeson, F. P. & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*, 241–293.

Wiesen (2020) Society for Police and Criminal Psychology Conference

68

---

---

---

---

---

---

---

---

## 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.

Wiesen (2020) Society for Police and Criminal Psychology Conference

69

---

---

---

---

---

---

---

---

## References

- Lievens, F., Ones, D. S., & Dilchert, S. (2009). Personality scale validities increase throughout medical school. *Journal of Applied Psychology, 94*, 1514-1535.



---

---

---

---

---

---

---

---

## References

- Ones, D. S. & Viswesvaran, C. (1998). Gender, age, and race differences on overt integrity tests: Results across four large-scale job applicant data sets. *Journal of Applied Psychology, 83*, 35-42.



---

---

---

---

---

---

---

---

## 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/>



---

---

---

---

---

---

---

---



## References

- Roth, P. L., BeVier, C. A., Switzer III, F. A. & 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

- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.



---

---

---

---

---

---

---

---

## 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

- Wee, S., Newman, D. A. & Joseph, D. L. (2014). More than *g*: Selection quality and adverse impact implications of considering second-stratum cognitive abilities. *Journal of Applied Psychology*, 99, 547-563.



---

---

---

---

---

---

---

---

## References

- Wiesen, J. P. (2018). *Master Tutorial: Tools to Increase Diversity, Utility, and Validity in Hiring Police Officers*. 2018 Annual SIOP Conference; Minneapolis, MN.



---

---

---

---

---

---

---

---