Tools to Increase Diversity, Utility, and Validity in Hiring Police Officers

Joel P. Wiesen, Ph.D.

Contact: **jw@jpwphd.com**

33rd Annual SIOP Conference
Chicago, IL, 4/20/2018

Print and Audio Links

- Conference proposal for this tutorial
- PowerPoints (pre-conference posted)
- Audio recording (yet to be posted)
- Links to related files
- **http://jpwphd.com/siop2018**

Questions

- Please hold questions until the end.
  – Much material to cover
- Review PowerPoints after the tutorial

Ultimate Goal

- Reinvigorate efforts to hire a diverse police officer (PO) workforce while substantially maintaining or improving the expected level of job performance (utility).
  - We have a moral, societal imperative to strive for diversity in hiring POs.

Tutorial Topics

- Appetizer

- Why search for new ways to select police officers (POs)

- Theoretical considerations (50% of time)

- Describe various tools (25%)

- Real-life examples (15%)

Appetizer

- Naiveté?

- Intellectual honesty?

- Do the tools really work?

- Balanced perspective

- Dual goals
  – Hire more minority POs
  – Improve expected job performance

Prevalent Wisdom

- *g* is the best test area: highest validity
- There is not much beyond *g*
- Don't dilute validity of *g* w/ low validity test
- Can select good employees with test of *g*
- Strong risk of increased *d* with 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
- Many false positive hires with a test of *g*
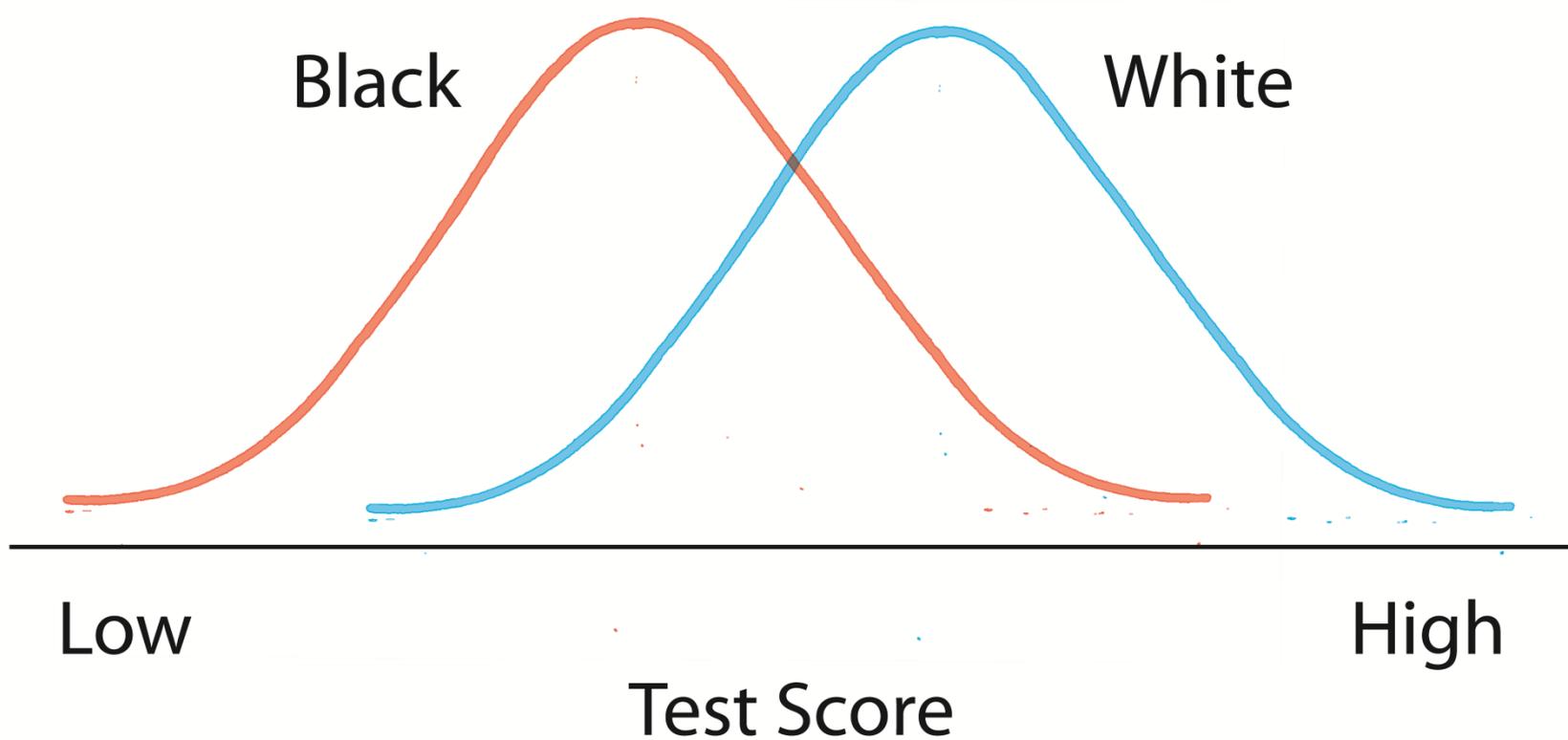- Composites usually have lower *d* than *g*

One Real-Life Example

- Large civil service agency

- Opened an application period for PO exam

- Not many minority applicants

- Decision: Extend the application period

Evaluate it Using I/O Methods

- Knowledge about d
- Knowledge about shape of distributions
- Knowledge about areas of normal curve

One Standard Dev. Difference

Normal Curve Analysis

- Adverse impact (AI) is a function of sample size as well as d

- Assume $d = 1$

- Use the Excel function NORMSDIST

AI for Various z-Score Cutoffs

z Score	Adverse Impact
-3	.98
-2	.86
-1	.59
0	.32
1	.14
2	.06
3	.02

Project Number of Hires

- Assumptions
- 10,000 applicants
  - 9,000 White
  - 1,000 Minority
- 500 openings
- Selection ratio = .05 (i.e., 5% or 1 in 20)
- 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

Result of Additional Recruitment

- Adverse impact expected to be worse

- Fewer minority hires expected

- Decision to extend the application period was **misguided**
  - Ineffective
  - Costly

Takeaways

- Analyze **quantitatively** before acting
- More applicants=AI & fewer minority hires
- Recruitment matters (big time)
- Try to recruit a higher proportion of minority applicants
- Try to recruit higher ability applicants
- Evaluate recruitment sources over time

Takeaways

- Consider trade-offs between investing in test development and in recruitment
  - Cost

  - Quality of hires

  - Number of minority POs hired

Tutorial Topics

- Appetizer
- Why search for new ways to select police officers (POs)
- Theoretical considerations
- Describe various tools
- Real-life examples

Why New Ways to Select POs?

- Social considerations

- Psychometric considerations

Social Considerations

- Our tests have adverse impact (AI)
- Many mixed munis. have mainly white PDs
  - e.g., Diversity on the Force (2015)
- Our cities are burning
- Our clients are being sued
- Our field is evolving, albeit slowly
- Psychologists' past support of **eugenics**

Psychometric Considerations

- *g* weakly predicts PO job performance
- *g* drives adverse impact (AI)
- Even low weight for *g* causes composite AI
  - Sackett & Ellingson (1997, Table 2)
- Other predictors have good r & smaller *d*
- Questions about fairness of our tests of *g*

Tutorial Topics

- Appetizer
- Why search for new ways to select police officers (POs)
- Theoretical considerations
- Describe various tools
- Real-life examples

Theoretical Considerations

- Main cause of adverse impact
- Do tests of *g* identify good POs?
- Level of validity of *g* for Police Officer
- Use utility or validity to select tests/KSAPs
- Are tests of *g* fair?

Main Cause of Adverse Impact

- Ranking candidates based on M/C tests of general mental ability (GMA), aka *g*
  - Cognitive ability
  - General intelligence
- One standard deviation difference (*d*) in mean scores for blacks and whites
  - Results in severe adverse impact on blacks

Main Cause of Adverse Impact

- Including g in a composite will cause AI
- Hard to get composite $< .5$ when g has $d = 1$
  – Sackett & Ellingson (1997, Table 3)
- Any use of g to rank, even in a composite, is likely to result in severe adverse impact
  – Sackett & Ellingson (1997, Table 2)

Do tests of *g* identify good POs?

- How do police managers view our tests?
- Psychometric analysis

Police Managers' View of Tests

- In favor:
  - Recruit many applicants to take test
  - Test is a fair way to identify good candidates to hire

- Against:
  - Tests of *g* ignore many important abilities
 - Hard to hire diverse a police force w/ *g* tests

Psychometric Analysis

- Do tests allow hiring of good POs?

- What is false positive rate?

- What is false negative rate?

- (Analyses of expected mean job performance yield basically the same conclusions as the simpler P/F analyses.)
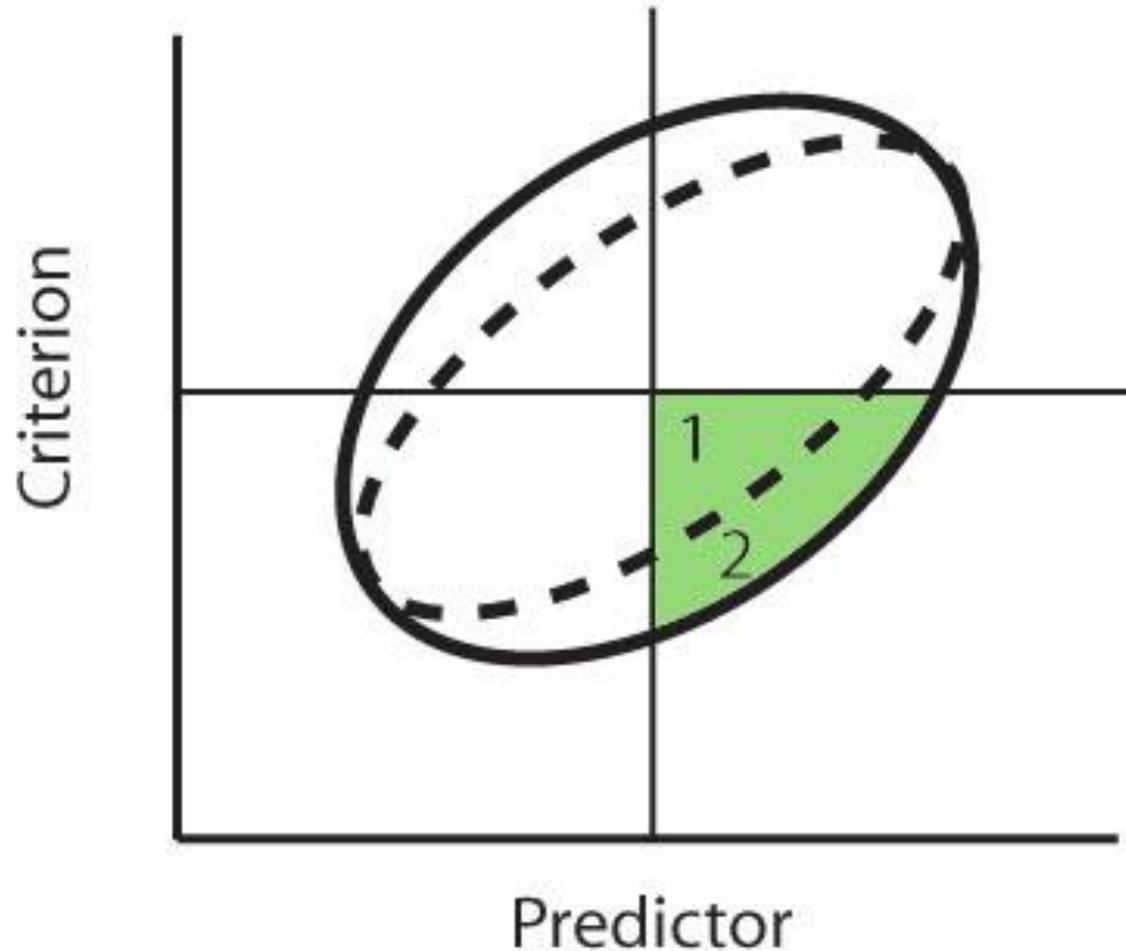
Definitions

- False Positive:

  A person cannot do the job but is hired.

- False Negative:

  A person could do the job but is not hired.

M/C Predictive Validity, r=.24
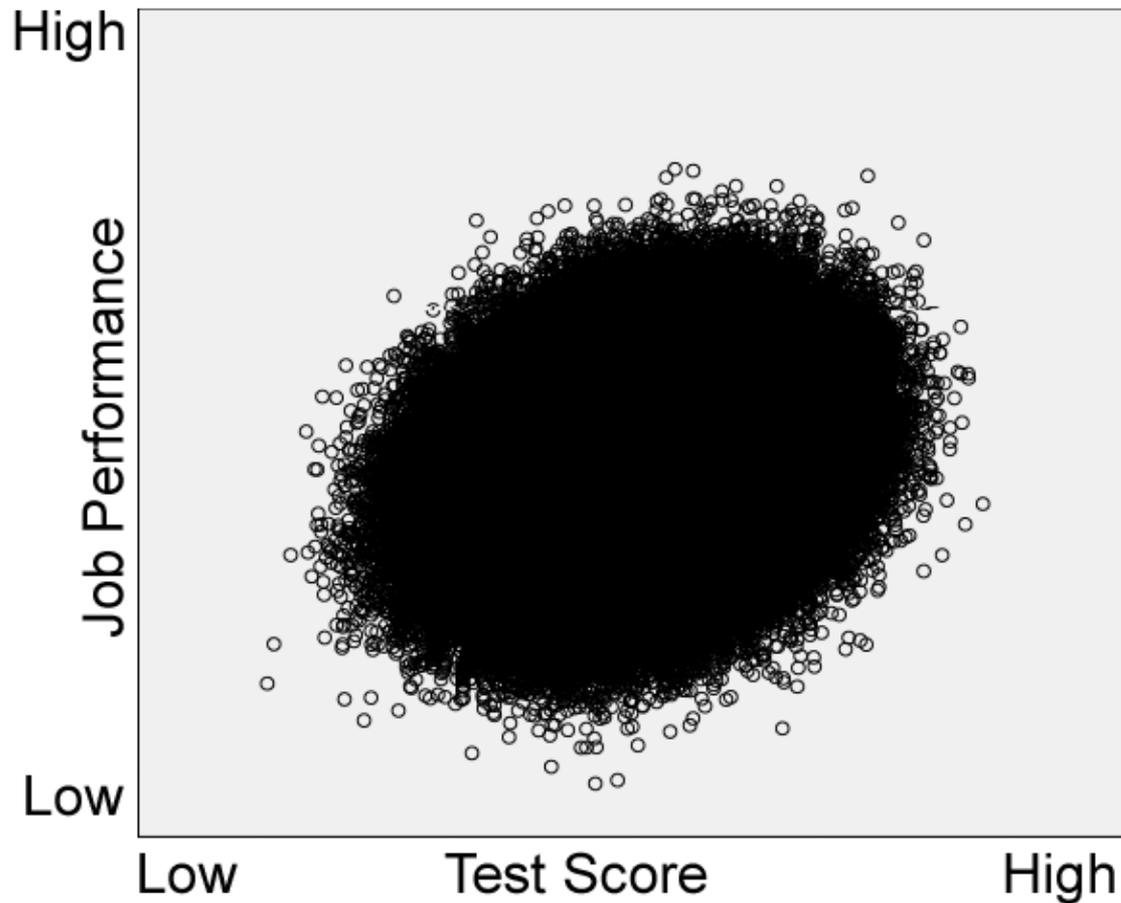
False Positives: Two Levels of r

Decisions, Right and Wrong

Textbook Expectancy Chart

Group	Chances of hires being successful (r=.7)
top 20%	90%
top 40%	80%
top 60%	70%
top 80%	60%
All	50%

(Based on Taylor & Russell, 1939, page 575; Q=.5)

Expectancy Chart, Q = .5

- Let's use job performance to index utility
- Look at r = .25 and r = .20

Expectancy Chart, Q = .5

Group	Chances of hires being successful (r=.25)	Chances of hires being successful (r=.20)
top 20%	64%	61%
top 40%	60%	58%
top 60%	56%	55%
top 80%	54%	53%
All	50%	50%

(Based on Taylor & Russell, 1939, page 575)

Expectancy Chart, Q = .5

- Utility of practically useful size (11+%)
- High proportion of false positives
  - 36% or 39% (considering *g* alone)
- Utility driven by SR as much as r
  - Within typical ranges of SR and r

Unmeasured Abilities

Let's assume there are untested KSAPs:

- Creative problem solving: 10% deficient

- Oral communication: 10% deficient

- Ability to get along w others: 10% deficient

- Conscientiousness: 10% deficient

- ~34% lack abilities not tested by M/C test

Revaluate False Positive Rate

- Expectancy chart: 61 to 64% true positives

- But 34% of these are deficient on non-g

- These abilities probably are independent

- So, reduce the 64% by 34% = 42%

- 42% true positives

Revaluate False Positive Rate

- Conclusion:
Most hires based on *g* are false positives
  - 58% false positives based on a typical test of *g*

What Happens with Higher Q?

- We hire better people

- Less room for improvement over chance
  - Cannot do much better than hiring randomly
  - Utility is lower

Expectancy Chart, Q = .9

Group	Chances of hires being successful (r=.25)	Chances of hires being successful (r=.20)
top 20%	95%	94%
top 40%	94%	93%
top 60%	93%	92%
top 80%	92%	91%
All	90%	90%

(Based on Taylor & Russell, 1939, page 577)

Compare Q= .5 and Q=.9

- Utility of r=.25, Q=.9 is 5% more true pos.
- Utility of r = .2, Q=.5 is 11% more true pos.
- Lower validity can have higher utility
- It depends on Q for the two areas tested
- In PD requiring college, Q for g may be high
- Q for a non-cognitive variable may be low

Some Takeaways

- Recruiting good applicants is **important**
- Tests of *g* result in many false positive hires
- Ignoring non-*g* areas inflates false positives
- Utility can be greater with lower validity test

How Valid is *g* for PO?

- r = .27 (corrected) for supervisor evaluations
- r = .62 (corrected) for academy grades
  - Aamodt (2004a)

Why is r for Academy Greater Than r for Job Performance?

- Reason to believe the higher value of r = .62
  - Job performance less reliably measured

Why is r for Academy Greater Than r for Job Performance?

- Reason to believe the lower value of r = .24
  - Common method (M/C) is unrelated to job
  - Both *g* test and academy grade are *g* loaded
  - Job perf. partly due to extra-individual factors
  - Job performance is partly due to non-*g* factors

Why is r for Academy Greater Than r for Job Performance?

- We assume that people who learn faster learn better.

  – But there is research to the contrary

- Slower learning can result in better retention and generalization

  – Bjork (2018, page 147, col 2, par 3), commenting on the idea introduced in his influential 1992 article in *Psychological Science*

Why is r for Academy Greater Than r for Job Performance?

- Validity of *g* decreases with time
- Validity of job knowledge increases w/ time
  - e.g., Farrell & McDaniel (2001)
- Validity of personality can increase w/ time
  - r =.18 to r=.45, year 1 to year 7 of med school
  - Lievens, Ones & Dilchert (2009).

Why is r for Academy Greater Than r for Job Performance?

- "Ability tests are commonly validated against narrow, cognitively loaded criteria"
  - Sackett, Shewach and Keiser (2017)

Focus on Utility, Not Validity

- Utility ≠ Validity
- Utility: "projected productivity gains … due to use of the selection procedure"
  - SIOP (2003, 4th edition, pages 49)
  - Let's ignore cost of testing and consider utility in terms of job performance
- Validity: "evidence and theory support ... proposed uses of … selection procedure"

Is Utility or Validity Primary?

- Profession seems to largely ignore utility
- Utility and validity are not identical
- A less valid test can have higher utility
- Selecting tests on utility may favor diversity
- Management interested in **utility**

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

Relationship Between U and V

- Validity drives utility
- Utility does not drive validity

What Drives Utility?

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

Some Takeways

- Lower r can have higher utility
  - Within the ranges of r that we often see
- It depends on the Q for the abilities tested
  - Q = the % of applicants who can do the job
- Selecting tests based on utility may favor diversity while improving job performance

APA: Unfairness is Serious

- "If ... excluding some components ... has a noticeable impact on selection rates for groups ... the intended interpretation of test scores ... would be **rendered invalid**."
  - AERA, APA, NCME (2014, page 21, col 1, par 1, emphasis added)
- So, the joint Standards says lack of fairness invalidates any indications of validity

Fairness of Our Tests of *g*

- Indications of fairness

- Indications of unfairness

Indications of Fairness

- Same regression line for various subgroups
- Same d for subjective and objective criteria
- Over predict for minorities

Indications of Unfairness

- *d* on test of g is larger than *d* on the job
- Criteria may be contaminated
- Relatively more minority false negatives

d for *g* Larger than *d* for Job

- d = 1.0 on tests of g
- d = 0.5 on job performance
- Is it explained by the regression formula?
    y = rx
- Assume r = .5
- We get .5 = .5(1.0)
- This seems to work.

d for *g* is Larger than *d* for Job

- But we selected based on a test of *g*
- So *d* should be less than 1
- *d* about .1 (if cut score set at +2 s.d. for W)
- Use the regression formula y = rx
- We get .5=.5(.1)
- This math does not work.

Criteria May Be Contaminated

- Tall people paid more than short
    - Judge & Cable (2004)
- Pretty people paid more than homely
    - Marlowe, Schneider & Nelson (1996)
- Men paid more than women
- Ethnic discrimination on the job
    - Umaña-Taylor (2016)

Reasons to Doubt Fairness

- There may be bias in supervisor ratings
- Aggressions affect job performance
- We predict unfair criterion accurately
- Indications of differential validity
  - e.g., Aguinis, Culpepper & Pierce (2016)
- If tests of g are not completely fair, try to limit their impact on selection

More Minority False Negatives

- "a given selection score … will often result in proportionately more false negative decisions in groups with lower mean test scores"
  - AERA, APA, NCME (1999, page 79)
- What this means: Among qualified applicants, the selection rate will be smaller for the lower scoring group.

Evaluate False Negatives

- How many false negatives?

- How different are the rates by ethnic group?

False Negative Scenario

- Assume:
  - Hire the top 1% of test takers
  - 10,000 test takers, 100 hires
  - 25% of W test takers can do the job
  - B-W $d = 1.0$
  - r = .24 (validity for PO, nationwide)
  - 25% of test takers are black

Outcomes

- Hire 2 black POs (out of 100 hires)

- Equity would be 25 black POs

- Adverse impact = .02 (very severe)

- 360 false negative black applicants

  – 360 blacks who can do the job but are not hired

- Hire 2 black POs but miss 360

Black Candidates

360 Could Do Job but Are Not Hired (False Negatives) and 2 Hired, Using Traditional M/C Test

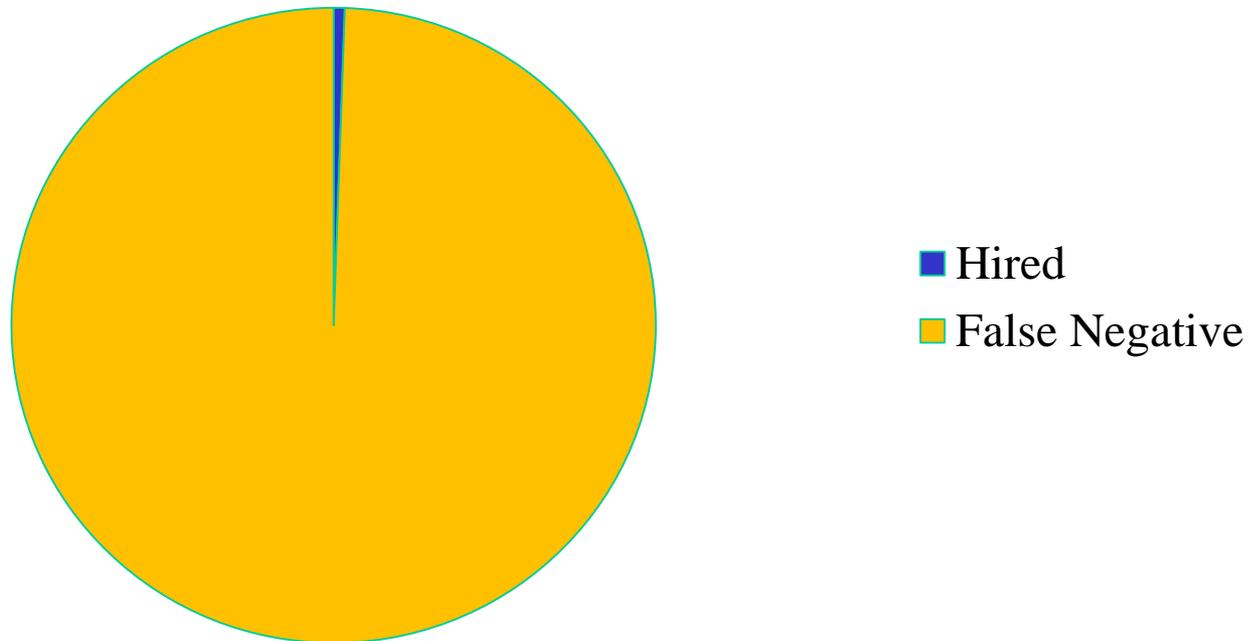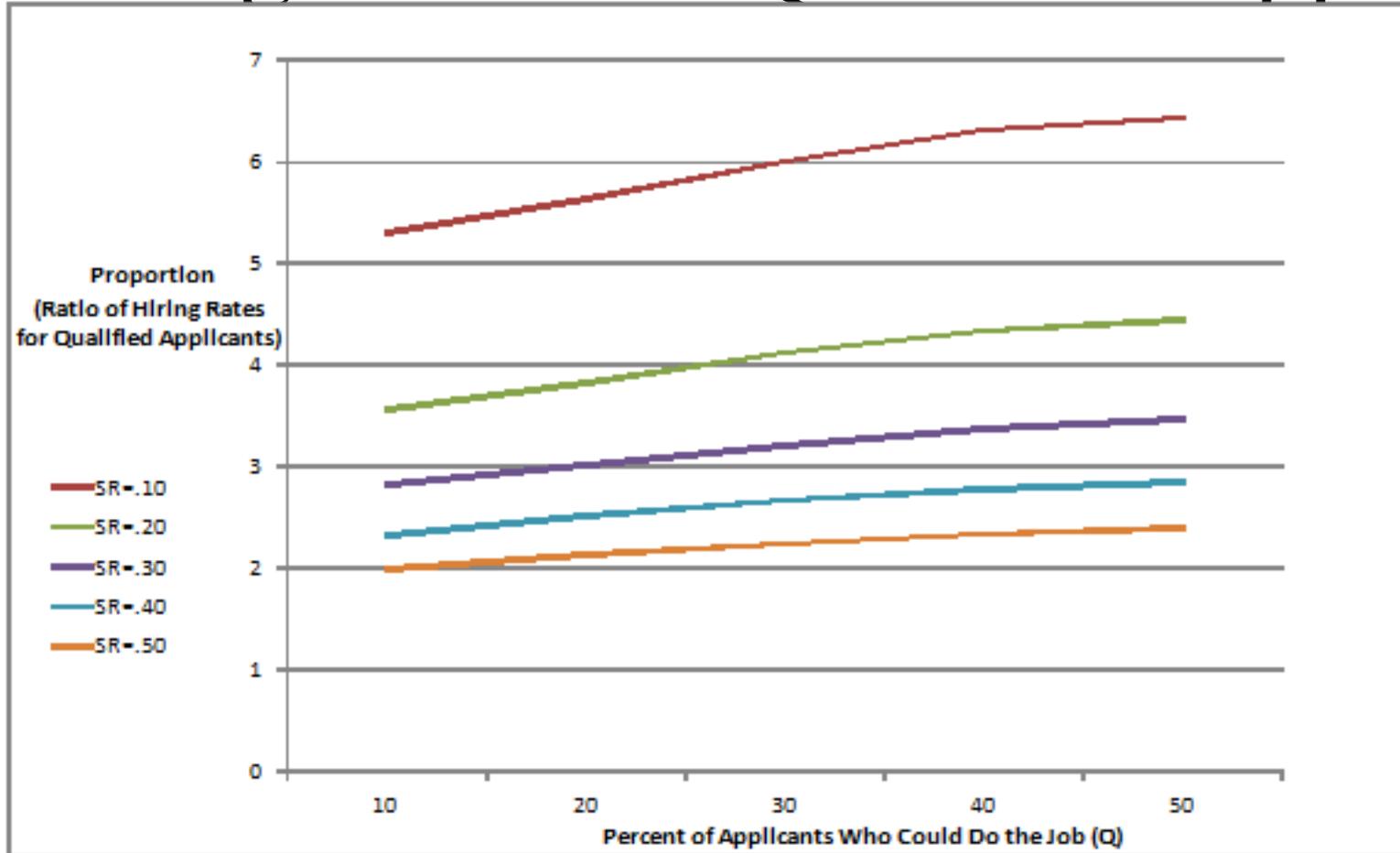Legend:
- Hired
- False Negative

Figure 2. W-B Disparities in Hiring Rates for Qualified Apps.

Some Takeaways

- There are reasons for disquiet about test fairness
  - Possible bias in criteria
  - Differential validity sometimes if found
  - Smaller proportion of qualified blacks hired

Tutorial Topics

- Appetizer
- Why search for new ways to select police officers (POs)
- Theoretical considerations
- Describe various tools
- Real-life examples

Fifteen Tools in Five Categories

- Administrative

- Use a traditional M/C test P/F

- Provide substitutes for the P/F M/C test

- Replace the traditional M/C test

- Rank based on job-related abilities with little, no, or **reverse** impact

Description of Tools

- Tool name

- What it is or how it works

- Is it legitimate?

  – Quote and source

- Practicalities (for some tools)

- Pros and cons

1. Empower the PD to Decide

- Provide managers with numeric projections
- Empower managers to make tradeoffs

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

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

Numeric Projections

- Project number of hires by protected class
- Project adverse impact
- Project utility
- Describe cost and timeline

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**	**Selection System 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

Police Officer Selection System Proposal Evaluation Form		
Topic	**Selection System Approach 1**	**Selection System Approach 2**
4. **Cost**		
Overall (for projected # of applicants)		
Cost adjustment +/- for more or fewer		

Proposal Comparison Form

Police Officer Selection System Proposal Evaluation Form		
Topic	**Selection System Approach 1**	**Selection System Approach 2**
5. Time to Implement Full System		
Days to develop selection system		
Days to implement selection system		

Proposal Comparison Form

Police Officer Selection System Proposal Evaluation Form		
Topic	**Selection System Approach 1**	**Selection System Approach 2**
6. **The basis for ranking candidates**		
Narrative entries		

Proposal Comparison Form

Police Officer Selection System Proposal Evaluation Form		
Topic	Selection System Approach 1	Selection System Approach 2
7. **Evaluation of Proposer**		
Past experience		
Expertise		
References		

Proposal Comparison Form

Police Officer Selection System Proposal Evaluation Form		
Topic	**Selection System Approach 1**	**Selection System Approach 2**
8. **Quality of Proposal**		
Understandable, logical, complete		
Believable projections, milestones		
Selection system practical		

1. Empower PD: Pros

- Focus decision making on likely results

- Hiring projections essential for managers

- Managers are responsible for the testing decisions, so they should make the decisions

1. Empower PD: Cons

- Projections somewhat difficult to make

- Hiring projections is risky for consultants

- AI ratios have relatively large std. devs.
  - Projections may be off

2: Use Short Application Period

- Limit the number of applications
- "The effects of group differences are greater as an organization becomes more selective."
  - Sackett & Ellingson (1997, page 511)
- Larger applicant pool = more selective
- Larger applicant pools will result in more severe adverse impact

2. Short Application Period: Pros

- Maintain validity

- There is no professional standard or principle concerning selection ratio (SR)

- Practical to implement

- Candidates accept it as fair

2. Short Application Period: Cons

- May lower the utility of selection procedure due to higher SR.

3. 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

3. Residency Preference: Pros

- Maintain use of traditional, valid predictors

- Historical precedent

- Practical to implement

- Candidates accept it as fair

3. Residency Preference: Cons

- May lower the utility of selection procedure

- Can be hard to determine residency

- Wealthier candidates can maintain two residences (i.e., lie more easily)

4. Use Test of *g* Pass-Fail (P/F)

- Really?
  - Does this fly in the face of all research?
- In support of ranking on *g* for PO
- In support of using *g* P/F for PO

In Support of Ranking Based on *g*

- *g* highly valid for jobs in general
  - Hunter & Hunter (1984)
- Linear relationship between *g* and job perf.
  - SIOP (2003, page 21)
- M/C *g* test inexpensive for large groups
- Takeaway: Always rank based on *g*
  - Adverse impact is unfortunate but unavoidable

In Support of Using *g* P/F for PO

- "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."
  – Sackett, Shewach, Keiser (2017)
- Assessment centers seem to have higher validity than tests of *g*, in general.
  – Why not rank on the test with highest validity?

In Support of Using *g* P/F for PO

- 25% of POs needed 2 year degree or more
  - 36% of PDs in cities of over 1,000,000
  - Reaves (2015, page 1, Fig. 7, and Table 7)
- If assume:

  r = .4 for *g* for the population

  40% of population have college degree
- Then validity of *g* for college grads = .23

In Support of Using *g* P/F for PO

- What of other 75% of smaller PDs
  - with no college degree requirements
- Low r for *g* for PO
- M/C tests of *g* are deficient
- Questions on fairness of tests of *g*
- Questions on fairness of criteria

Low r for Police Officer

- Hunter & Hunter (1984), r = .38 for PO
  - One of the lowest r's for a specific occupation
- Aamodt (2004a) meta-analysis
  - r = .27 (over corrected for practical purposes)
  - r = .24 more realistic for practical purposes

All PDs Require a HS Degree

- 75% of POs work for a PD with no college requirement
- All PDs require at least HS

What is the Mean *g* for POs?

- Mean IQ for police officers = 104
  - Aamodt (2004b, based on total of 4,061 POs)
- What is mean IQ for high school grads?

M/C tests of *g* are Deficient

- Creative problem solving

- Correlation between *g* and leadership

- How other fields of psychology view *g*

- There are newer tests of intelligence

Creative Problem Solving

- Cannot fully measure creativity with a M/C test
  - Only open-ended questions allow for original responses
- r = .07 to .29 for creativity and g
  - e.g., Kim (2005), Sternberg (2006, Tables 9, 11.1, 11.2)

Correlation of *g* and Leadership

- Intelligence is the ability to deal with the demands of the real world in order to achieve success

- Leading others is a demand of the real world

- r = .19 if *g* measured with M/C

- r = .6 if *g* measured by ratings others make
  - Judge, Colbert & Ilies, 2004 (Table 2)

Broader Psychological View

- "Most current large-scale testing is not well suited to [measure] higher order thinking skills, collaboration, information literacy, etc."

  – Dolan, Goodman, Strain-Seymour, Adams & Sethuraman (2011, page 5)

Broader Psychological View

- "Traditional tests of intelligence are not good proxies for rational thinking skills."
  - Stanovich, West & Toplak (2012)

Broader Psychological View

- "Standard conventional tests only assess a narrow sampling of the abilities required for success in school and in life."
  - Sternberg (2015)
  - His are ideas worth exploring

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*

- There are facets to *g*

- Some facets have smaller *d*s

- Some composites of *g* yield less adverse impact and more diverse hiring

- Facets of *g* are not equally valid for various different jobs
  - e.g., Wee, Newman & Joseph (2014)

Questions on *g* Test Fairness

- This was covered earlier

In Support of Using *g* P/F for PO

- Why rank candidates based on a low validity test (of *g*) that almost guarantees severe adverse impact?

- Especially since there are unresolved issues of test fairness

- Especially since there are other valid tests/tools that could be used with good utility

4. Use Test of *g* P/F: Pros

- Assure a level of *g* similar to actual POs

- Allows for more diversity in hiring

- Ranking on other KSAPs can increase utility and validity

4. Use Test of *g* P/F: Cons

- Does not assure maximum level of *g*

- Validity and utility depend on the availability of other tests

- How to respond to "lowering standards"
  – Measure all the important KSAPs, including *g*

  – Changing standards

  – Goal posts not closer; repositioning goal posts

5. Alternatives to a P/F Test of *g*

- Honorable 3 year military discharge
- Associate's or higher college degree
- Sufficient high school rank or GPA
- Oral exam
- Essay exam

5. Alternative to P/F *g* Test: Pros

- Measure/assure *g* with less adverse impact

5. Alternative to P/F *g* Test: Cons

- Validity of an alternative may be lower
- Some alternatives may not be practical
  - Some high schools may not release HS rank
- Some alternatives are more expensive
  - Oral exam, essay exam
- Some alternatives may have adverse impact
  - College degree (check local demographics)

How Rank Candidates if Not on *g*

- Several tools to do this

- Some tools can be used together, perhaps in composites

6. Rank Candidates Based on HS

- High school rank
  - Not all schools release rank
- High school GPA
  - r = .91 between GPA standardized within school and not
  - Sternberg (2006, page 331, col. 1, par. 3)

Validity of Schooling

- 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)

6. Rank on High School: Pros

- Levels the playing field when students from poor, primarily minority inner-city schools compete with white students from rich suburban schools

- School grades and rank may be easily available

- Likely good candidate acceptance

6. Rank on High School: Cons

- Need to provide for the less conventional
  - Home schooled
  - GED
  - Ungraded schools

7. Rank Based on a Structured Oral Exam

- Highly valid
  - the most valid, r = .57
  - Aamodt (2016, Table 5.2, page 194)
- In the past 20 years, the B-W d for structured interviews has been zero
  - Levashina, Hartwell, Morgeson & Campion (2014, Table 3, page 254)

Managers Like Structured Interviews

- One study found the interview to be the most trusted by police managers of the 10 selection measures administered
  - De Soete, Lievens, Oostrom & Westerveld (2013, Table 2).

One Study of *d* for PO

- Structured interview for PO

- N=1,334 applicants

- *d* slightly below zero
  - McFarland, Ryan, Sacco & Kriska (2004, Table 2)

7. Rank on Structured Int.: Pros

- Validity may be higher than a test of g
- Content validity higher than test of g
- Can measure much more than g
  - e.g., Cascio & Aguinis (2011), p. 268, par. 4
- Likely high candidate acceptance
  - Depends on the questions and interviewers
- Expect low or zero d

7. Rank on Structured Int.: Cons

- Difficult to maintain security of questions

- Lower reliability than a M/C test

- Inter-panel differences must be addressed

- Relatively costly to administer

- Candidate demand for transparency may threaten security

8. Rank Based on Each Candidate's Highest Score

- Greatest Strength Method (GSM)

- Set passing score on each KSAP

- Use highest score to rank candidates
  - Highest z score

Support for Highest Score Method

- Large Monte Carlo studies using plausible scenarios

- Compared GSM to g alone and to several weighted composites

Support for Highest Score Method

- Modeled job of Firefighter
- Minority and female applicants
- Varied SR and % minority applicants
- Looked at r, d, AI, EMJP
  - Expected mean job performance (EMJP)
  - Wiesen & Aquinis (2010)

GSM vs *g*

Method	Validity	AI EEO	AI Gender	MJP
GMA	**0.51**	**0.31**	**1.00**	**0.45**
GSM	0.52	0.63	0.66	0.57
Random	0.00	1.00	1.00	-0.27

GSM showed:

equal validity

less severe ethnic AI

higher mean job performance

8. Rank on Highest Score: Pros

- Monte Carlo studies support the GSM
  - Maintain validity
  - Improve mean job performance
- Reduced AI on all protected groups
- Similar to teacher dropping the lowest grade
  - No student outcry about lost information

8. Rank on Highest Score: Cons

- Only Monte Carlo research

- Only conference papers

- Large standard deviations of AI ratios
    - Your result may be far from the mean

- Candidates who score high on all KSAPs may not like this approach

- Comparison is to *g* alone, not to all KSAPs

Forming Composites for Ranking

- Don't include *g*

- But if you do, don't over weight *g*

- The next tools are variables/tests to consider using in a composite to rank candidates

Easy to Over Weight *g*

- Assume weighting based on content validity
- Assume relatively large Q for *g*
  - e.g., college degree required
- z-score weighting will spread out the restricted range of *g* scores
- In effect, this will over weight *g*
  - In comparison to a factor with lower Q

Do Composites Dilute Validity?

- Evaluate this mathematically, analytically

Concern: Dilute the Validity of *g*

- Here are two responses to this concern
- If *g* is most valid, will adding a low r test dilute that validity?
  - No: validity tends to sum, not average
- We might best focus on utility, not validity

Use Psychometric Formulas

- Validity of the sum of two tests
- *d* for the sum of two tests
- Multiple correlation with two predictors
- Sum of two weighted measures with a third

Validity of the Sum of 2 Tests

- Correlation of a sum of two weighted measures with a third measure

$$r_{c(ws)} = \frac{w_1 r_{c1} \sigma_1 + w_2 r_{c2} \sigma_2}{\sqrt{w^2_1 \sigma^2_1 + w^2_2 \sigma^2_2 + 2 r_{12} w_1 \sigma_1 w_2 \sigma_2}}$$

(Guilford, 1965, page 427, formula 16.25)

Sacket & Ellingson, Formula 3*

$$d = \frac{\sum_{i=1}^{k} w_i d_i}{\sqrt{\sum_{i=1}^{k} w_i^2 + 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} w_i w_j r_{ij}}}$$

* Corrected last term in denominator; typo in 1997 paper

Predictive Formulas

- Multiple correlation with two predictors
  – Formula 16.1, page 394
  – Guilford (1965)

- Sum of two weighted measures with a third
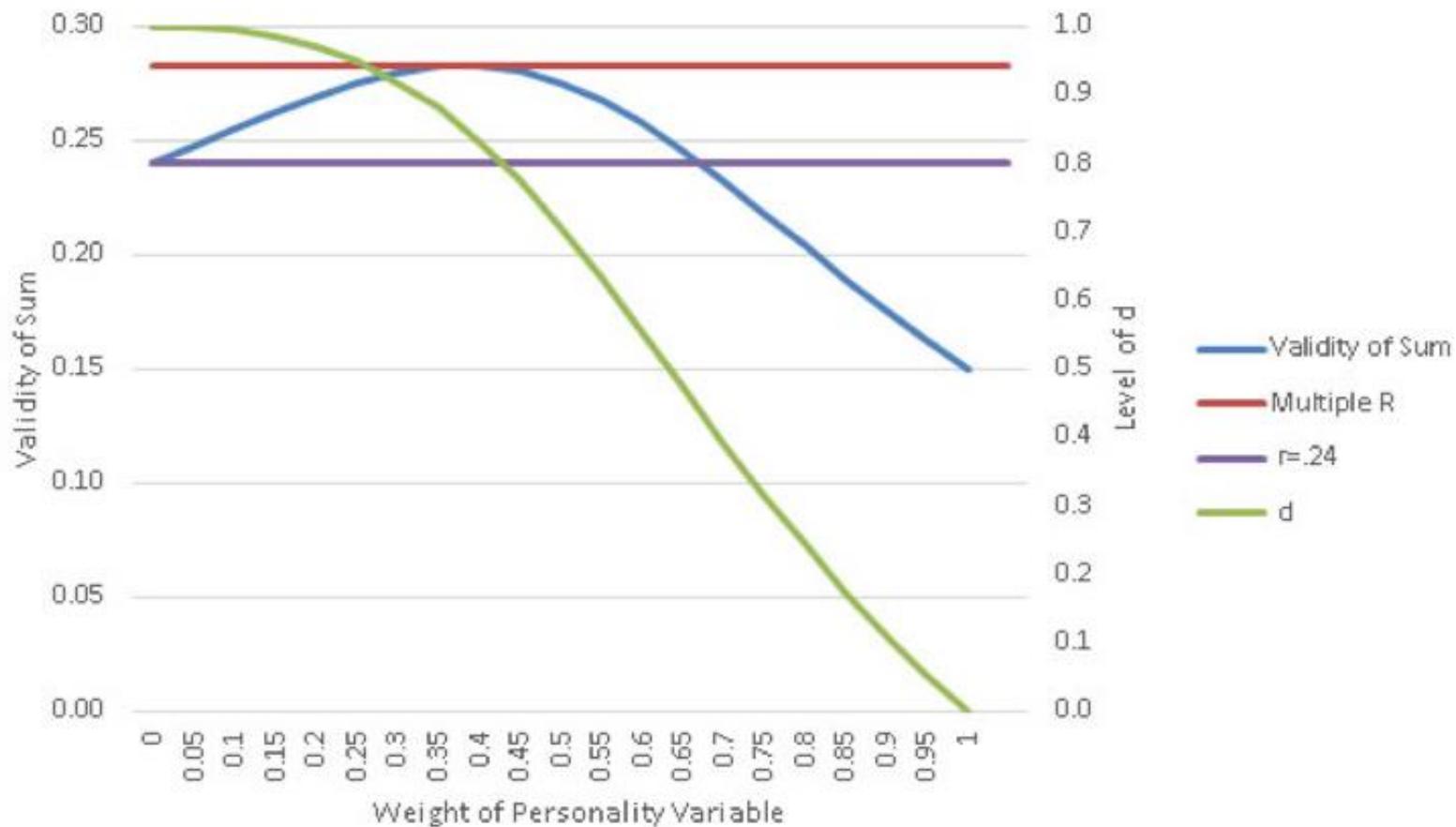 – Formula 16.25, page 427
  – Guilford (1965)

Predictive Formulas

- Some more detail here:

    http://jpwPhD.com/siop2018

Composites of Two Tests

- An example
- Assume tests with r = .24 and r = .15
  – e.g., *g* and a personality factor, uncorrelated
- Assume *d*s of 1 and zero, respectively
- When form a composite, what happens to:
  – Validity
  – *d*

Figure 1. Validity of Sum and *d* by Weight of Personality Variable

9. Rank Using Video Testing

- Evidence of increased validity and lower d
  - Chan & Schmitt (1997)
  - Lievens & Sackett (2006)

9. Rank Using Video Test: Pros

- Likely high user acceptance

- Can have good face validity

- Can assess many different KSAPs
  - Including and beyond g

9. Rank Using Video Test: Cons

- Lower reliability than tests of *g*
- Staging and recording actors is expensive
  - Salaries, sets, costumes, camera equipment, etc.
- Modifying such video tests is problematic
- Less expensive avatars are less realistic
  - Cartoon-like
  - Unknown equivalence to recordings of actors

10: Rank Using Memory of Faces

- Measure face recognition and memory
- Use faces that mirror the community
- Use faces that mirror the offenders

Expect **Reverse** Impact

- Remembering and identifying minority faces is easier for members of that minority group
  - e.g., Levin (2000)
- I am collecting data on this as we speak.

Validity

- I expect face memory/recognition would be supported by content validity
  - Likely related to various job tasks
  - Recognizing perps
  - Recognizing residents

10. Rank w/ Face Memory: Pros

- Expect content validity, esp. for some tasks

- Expect **reverse** impact (against whites)

- Candidate acceptance likely to be high

- Can be implemented in print form or online

10. Rank w/ Face Memory: Cons

- Largely untested
  - No existing tests for large groups
  - No criterion-related validity studies

11. Rank Using Physical Fitness

- PO is a physically demanding job
  - e.g., Shetterly & Krishnamoorthy (2008)
- Criterion-related validity (many rs in .20s)
  - Weiner (1994)

11. Rank w/ Physical Fitness: Pros

- Increase utility

- Increase mean job performance

- Candidate acceptance likely to be high

11. Rank w/ Physical Fitness: Cons

- Question of 1991 CRA: gender standards
- Will have adverse impact on women if the focus is on strength rather than fitness
- Gender standards may be a legal liability

12. Rank Using Test of Prejudice

- Ideal: A police force without any prejudice

- Can we measure prejudice?

Implicit Association Test (IAT)

- Based on reaction time and cognitive choice

- React to ethnic group faces

- React to emotionally laden or stereotypical objects or words

- Can try it on the web: http://implicit.harvard.edu

IAT Validity

- Modest validity, perhaps "poor validity"
  – Oswald et al. (2013)
- Yet even that study reported meta-analyses showing IAT validities of .19 for person perception and .14 for microbehavior (the two categories of studies with the largest N's in their Table 3, excluding EEG studies).

Expect **Reverse** Impact

- Minority group members tend to be less biased than white people against members of their own minority group
  - Axt, Ebersole & Nosek (2014, Table 1)

12. Rank Using Prejudice: Pros

- Expect **reverse** impact
- Add to predictive power of other tests
- Low cost (automated administration)
- Important characteristic
- Hard to coach for the test

12. Rank Using Prejudice: Cons

- Controversial type of test
- Validity for PO selection not yet researched
  - Would require huge N (500?, 1,000?)

13. Rank Using Other Personality Measures

- Many big-5 and smaller facet personality factors to choose from

- Some examples follow

Some Other Personality Areas

- Personality factors (e.g., conscientiousness) and facets with r's in the .15 to .20 range and with small or zero d's
    - e.g., e.g., Hough & Johnson (2013)
- Extroversion and emotional stability
    - Ployhart & Holtz (2008, Table 1)

Some Other Personality Areas

- Certain personality facets: achievement orientation, conscientiousness, surgency/dominance/potency facet of extroversion, and adjustment
  - Hough, Oswald & Ployhart (2001, Table 2)
- Effort (care and persistence)
  - r = .28-.33 with college GPA
  - Briley, Domiteaux & Tucker-Drob (2014)

Integrity Tests

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

Personality Validity Tends to Sum

- Validities for uncorrelated variables sum

- Personality has low correlation with g

- Personality validities should tend to sum
  - Schmitt (2014)

13. Rank Candidates Using Other Personality Attributes: Pros

- Will improve validity and utility by expanding job-related abilities tested

- Will tend to have little AI

13. Rank Candidates Using Other Personality Attributes: Cons

- Faking by applicants is an issue

- Candidate concern for lack of transparency

- Candidate demand for transparency may threaten security

14. Rank Using Creativity and Problem Solving

- Creativity to reduce AI in college admission
  - Kaufman (2010)

- Oral story telling and SAT correlate equally with college GPA, but with smaller d
  - r = .29 and r = .26 or .28 for story telling and SAT V and M, respectively
  - d = .14 vs .74 and .67
  - Sternberg (2006, Tables 11.1, 15)

Higher Validity, Lower AI

- Considering creativity and practical intelligence increased validity and decreased adverse impact
  - Sternberg (2015)

14. Rank w/ Creative PS: Pros

- Enhanced validity for college grades
- Reduce adverse impact
  - Sternberg (2015)
- Increase validity by decreasing predictor deficiency
- There are published tests of creativity

14. Rank w/ Creative PS: Cons

- Hand grading is labor intensive (expensive)

15. Focus on Utility, Not Validity

- Design selection systems based on utility
- Selection systems designed based on U may yield higher job performance and less AI

15. Focus on Utility: Pros

- Better job performance
- Less severe adverse impact
- Utility discussed in the Principles
  - SIOP (2003, 4th edition, pages 48-49)
  - SIOP (2017, draft 3-7-17 for review)
    - Little (no?) change to section on utility

15. Focus on Utility: Cons

- The profession has less clarity on how to use utility to design a selection system

- Focus on utility does not address the emphasis on validity in the *Uniform Guidelines on Employee Selection Procedures*

Tutorial Topics

- Appetizer
- Why search for new ways to select police officers (POs)
- Theoretical considerations
- Describe various tools
- Real-life examples

Real-Life Examples

- Police

- Fire

- Military

Bridgeport CT PD, 2014

- Mayor wanted to hire diverse class, validly

- Consultant proposed: average M/C and personality test to reduce impact of M/C

- Minority union asked my opinion of this proposal

- Opinion: City will hire few minority POs

- City asked for my recommendations

Tools Bridgeport PD Used

- M/C exam, as a qualifier

- Ranked candidates on a structured oral exam

- Additional residency points

Bridgeport Outcomes

- "... 61% are … minorities or women … a number we have never had before." [46% minority appointees]

- "We couldn't have been prouder of this process. This did not happen on its own; we made significant changes to the process."
Source: Only In Bridgeport (2015)

Miami, FL PD

- Consent decree for PO from 1977-2016
- Maximum of 500 applicants, 3 times/year
  - Occasionally the cap is 750 or 1,000 applicants
- State qualifying exam (Basic Ability Test)
  - About 12% fail the Florida BAT
- No municipal exam
- No ranking, passing the screening are hired

Miami, FL PD (continued)

- State "Basic Ability Test"
  - Fleishman areas
  - Low passing score?
- Intensive recruitment, incl. community gps.
- Physical ability test, gender and age normed
  - Developed with DOJ involvement
- Stringent background check

Miami, FL PD (continued)

- If there is need to rank applicants, Miami uses a structured oral exam that is not cognitively focused.

Los Angeles PD

- No M/C exam

- Rank based on score on the Personal Qualifications Essay (PQE)

- Detailed background check (P/F)

- Interview (P/F)

- Physical (P/F)

- Used for 10+ years, but now changing

LAPD PQE

- Essay exam administered in person

- Scored based on:
  - Written communication skill
  - Judgment and decision-making
  - Behavioral flexibility

- Candidates ranked based on the PQE

Columbus, OH, PD & FD

- P/F M/C cognitive ability test
- P/F writing exercise
  - Score on information analysis, writing skill
- P/F physical fitness test
- Rank on video test of problem solving
  - Still photos with verbal enactments of scenarios
  - Candidates respond orally
  - Score problem sensing/resolution, interpersonal
  - Columbus Civil Service Commission (2014)

Oklahoma City FD, mid-1990's

- FD did not hire any black FFs for several exams
- FD and minority union very concerned

Oklahoma City, OK, FD, 1990's

- P/F M/C cognitive ability test
- P/F 40 hour, First Responder course
- P/F written workstyle test
- P/F background check
- P/F physical ability test
- P/F medical
- Ranked oral exam

OCFD Outcomes

- Hired 7 black firefighters (48 white)

- Average grade in the training academy similar to previous classes

- Dropout rate from the training academy was a little lower than previous years

OCFD Outcomes (continued)

- Average quarterly recruit evaluations were high

- All the new firefighters became certified EMTs after being hired

- This selection system was used again in 1996 with similar positive results

Army Assessment of Background and Life Experience (ABLE)

- Personality: motivation and attitude

- Achievement, Adjustment, Agreeableness,

- Dependability, Leadership, and Physical Conditioning

- Predicts attrition and performance, independently of the ASVAB/AFQT
  - Wenger (2010, page 4)

Army Assessment Individual Motivation (AIM)

- Forced-choice approach employed
  - Less susceptible to faking
  - Behavioral questions
- Same areas as ABLE
- Predicts attrition and performance, independently of the ASVAB/AFQT
  - Wenger (2010, Tables 1 and 2, Figure 7)

Tier Two Attrition Screen, TTAS

- Personality measure

- Measures adaptability

- Predicts military retention

- Smaller *d* for gender and race/ethnicity
  - White, Rumsey, Mullins, Nye & LaPort (2014, page 147)

Let's Recap

Problem is Clear

- In many (all?) circumstances, we can reduce *d* and increase validity by including a personality test with a test of GMA

- But reduced *d* still causes adverse impact

Requirements for a Solution

- Find ways to maintain high validity and utility while reducing impact of test of *g*

How Do the Tools Work?

- Use what we know to predict d and AI
  - Formulas (e.g., Ghiselli et al., 1981)
- Some KSAPs have small or even negative d
- Focus on utility over validity
- Utility driven more by recruitment than validity

What Tools Are Not

- No smoke and mirrors
- No chicanery

What is Professionally Acceptable?

- Local criterion-related validity study
- Meta-analysis for the job of PO
- Meta-analysis for jobs in general
- Content validity for job simulation test
- Content validity for "constructs"
- Representative sample of job needed?

Professionally Acceptable?

- Some, all or most critical aspects of the job?

- Majority of KSAPs required or ideal?

- What level of validity?

- What level of utility?

- Relaxed criteria if no adverse impact?

- More stringent criteria if adverse impact?

- Relaxed criteria if **reverse** impact?

If We Are Too Conservative

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

Use of Novel Tests

- Can I use a test that is novel?

- Can I use a test that is controversial?

- Can I choose not to use a new promising test?

Review Learning Objectives

- Four objectives stated to get APA CE credit

First Learning Objective

- Describe the conditions under which a relatively low validity test (e.g., r = .15) is expected to have higher utility than a higher validity test (e.g., r = .25).

First Objective Takeaway

- Conditions: Lower r test has higher Q

- Utility depends on r, SR, and Q
  - Q is the quality of the applicants

- If all applicants have high ability, a test has low utility
  - e.g., college degree as an entrance requirement

- Often there is more candidate variability in personality than in cognitive ability

Second Learning Objective

- Describe at least 2 test areas that generally have **reverse** impact on minorities and 2 test modes with low d.

Second Objective Takeaway

- Test areas
  - Minority face recognition
  - Prejudice
  - Certain personality facets: achievement orientation of conscientiousness, surgency/dominance/potency facet of extroversion, and adjustment
    - Hough, Oswald & Ployhart (2001, Table 2)

Second Objective Takeaway

- Test modes
  - Oral/Video
  - Constructed response
  - IAT

Third Learning Objective

- Describe the pros and cons of using tests of *g* on a pass-fail (P/F) basis.

Third Objective Takeaway

- Pros:
  - More diversity
  - Use/design new, better selection system with higher utility
- Cons:
  - There is a danger of losing validity
  - There is a danger of losing utility

Fourth Learning Objective

- Explain why ranking even in part on a traditional test of *g* generally results in adverse impact on minority candidates.

Fourth Objective Takeaway

- *g* almost always shows large *d*
- Composites with *g* also have sizable *d*
- Even modest *d* can result in severe AI

In Closing

- Goal:

  – Reinvigorate search for alternative selection procedures to maintain or improve job performance and achieve diversity.

Summary and Conclusions

- *g* may not be the most valid test for PO

- *g* may not have highest **utility** for PO

- Use *g* cautiously (e.g., P/F)
    - Do not let *g* drive severe adverse impact

- Ranking based on *g* will create adverse imp.
    - Even using a composite w/ low weight for *g*

- Use valid alternatives for tests of *g*

Summary and Conclusions

- Evaluate utility separate from validity
- Consider utility when choosing tests
- Recruitment often more crucial than testing
- *g* tests can have high false positive rates
  - Due to modest validity

Summary and Conclusions

- Use a content validity approach
  - Due to scarcity of criterion-related research
- Test many job-related abilities
  - Face memory, creative problem solving, conscientiousness, energy, prejudice, etc.
- Rank based on valid tests with d near zero
  - Structured oral exam, etc.

Do We Face a Moral Issue?

- Psychologists in forefront of the **eugenics** movement and Immigration Act of 1923
- Limit entry to the USA by country quotas
- Favored England and Western Europe
- Southern and Eastern Europe not favored
- Asians/Jews thought to be low IQ, undesir.
- How could psychologists be so wrong?

Do We Face a Moral Issue?

- Are I/O psychologists making errors today similar to those made by the psychologists who supported eugenics and restrictive immigration?

More Information

- Wiesen, J. P. (2016, 2017a, 2017b)

Whova This Session

- Add session to your agenda (for head count)
- Rate the session
- Use Comments for questions & feedback
- Like the session if you liked it

Links

- Link: **http://jpwphd.com/siop2018**

- More information on tools

- Some selection-related formulas

- PO Selection Proposal Evaluation Form

- Email: **jw@jpwphd.com**

- Telephone: **(617) 244-8859 (land/no text)**

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

- Aguinis, H., Culpepper, S.A., & Pierce, C.A. (2016). Differential prediction generalization in college admissions testing. *Journal of Educational Psychology, 108*, 1045-1059.

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

- Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2014). The rules of implicit evaluation by race, religion, and age. *Psychological Science, 25,* 1804-1815.

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

- Bjork, R. A. (2018). Being suspicious of the sense of ease and undeterred by the sense of difficulty: Looking back at Schmidt and Bjork (1992). *Perspectives on Psychological Science, 13*, 146–148.

References

- Briley, D. A., Domiteaux, M., & Tucker-Drob, E. M. (2014). Achievement-relevant personality: Relations with the Big Five and validation of an efficient instrument. *Learning and Individual Differences, 32,* 26-39.

References

- Cascio, W. F. & Aguinis, H. (2011) *Applied Psychology in Human Resource Management* (7th ed.) Saddle River, NJ: Prentice Hall.

References

- Chan, D. & Schmitt, N. (1997). Video-Based Versus Paper-and-Pencil Method of Assessment in Situational Judgment Tests: Subgroup Differences in Test Performance and Face Validity Perceptions. *Journal of Applied Psychology, 82,* 143-159.

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

- De Soete, B., Lievens, F., Oostrom, J. & Westerveld, L. (2013). Alternative Predictors for Dealing with the Diversity–Validity Dilemma in Personnel Selection: The constructed response multimedia test. *International Journal of Selection and Assessment, 21,* 239-250.

References

- Diversity on the Force: *Where Police Don't Mirror Communities; A Governing Special Report* (September, 2015). Washington: Governing.

References

- Dolan, R. P., Goodman, J., Strain-Seymour, E., Adams, J. & Sethuraman, S. (2011). *Cognitive Lab Evaluation of Innovative Items in Mathematics and English Language Arts Assessment of Elementary, Middle, and High School Students; Research Report*. San Antonio, TX: Pearson.

References

- Equal Employment Opportunity Commission. (1978). Uniform guidelines on employee selection procedures. *Federal register, 43,* 38295-38309.

References

- Farrell, J. N., & McDaniel, M. A. (2001). The stability of validity coefficients over time: Ackerman's (1988) model and the general aptitude battery. *Journal of Applied Psychology, 86*, 60–79.

References

- Ghiselli, E. E., Campbell, J. P. & Zedeck, S. (1981). *Measurement Theory for the Behavioral Sciences*. San Francisco: W. H. Freeman.

- Guilford, J. P. (1965). *Fundamental Statistics in Psychology and Education (4th ed.)* New York: McGraw-Hill.

References

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

- Judge, T.A. & Cable, D.M. (2004). The Effect of physical height on workplace success and income. *Journal of Applied Psychology, 89,* 428-441.

References

- Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology, 89*, 542-552.

References

- Hough, L. M., & Johnson, J. W. (2013). Use and importance of personality variables in work settings. In I. B. Weiner (Ed.-in-Chief) & N. Schmitt & S. Highhouse (Vol. Eds.), *Handbook of psychology: Vol. 12. Industrial and organizational psychology* (pp. 211-243). New York: Wiley.

References

- Hough, L.M., Oswald, F.L. & Ployhart, R.E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9*, 152-194.

References

- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,*72-98.

- Kaufman, J. C. (2010). Using Creativity to Reduce Ethnic Bias in College Admissions. *Review of General Psychology, 14*, 189–203.

References

- Kim, K. H. (2005). Can Only Intelligent People Be Creative?; A Meta-Analysis. *The Journal of Secondary Gifted Education, 16,* 57–66.

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.

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.

References

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

References

- Lievens, F. & Sackett, P. R. (2006). Video-Based Versus Written Situational Judgment Tests: A Comparison in Terms of Predictive Validity. *Journal of Applied Psychology, 91*, 1181–1188.

References

- Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology, 81*, 11-21.

References

- McFarland, L. A., Ryan, A. M., Sacco, J. M., & Kriska, S. D. (2004). Examination of structured interview ratings across time: the effects of applicant race, rater race, and panel composition. *Journal of Management, 30, 435-452.*

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

- Oswald, F. L., Mitchell, G. Blanton, H., Jaccard, J. & Tetlock, P. E. (2013). Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies. *Journal of Personality and Social Psychology, 105*, 171–192.

References

- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*, 153-172.

References

- Reaves, B. (2015, May). *Local Police Departments, 2013: Personnel, Policies, and Practices*. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, NCJ 248677. Downloaded December 10, 2017 from https://www.bjs.gov/content/pub/pdf/lpd13p pp.pdf.

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

- Schmitt, N. (2014). Personality and Cognitive Ability as Predictors of Effective Performance at Work. *Annual Review Organizational Psychology and Organizational Behavior, 1*, 45-65.

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

- Shetterly, D. R. & Krishnamoorthy, A. (2008) Job Characteristics of Officers and Agents: Results of a National Job Analysis. *Public Personnel Management, 37*, 111-131.

References

- Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the Validation and Use of Personnel Selection Procedures (4th ed.)* Bowling Green, OH: Author.

References

- Society for Industrial and Organizational Psychology, Inc. (2017). *The Principles for the Validation and Use of Personnel Selection Procedures* (3-7-17_for_review). Bowling Green, OH: Author

References

- Stanovich, K. E., West, R. F., & Toplak, M. E. (2012). Judgment and decision making in adolescence: Separating intelligence from rationality. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), *The adolescent brain: Learning, reasoning, and decision making* (pp. 337-378). Washington, DC: American Psych. Assn.

References

- Sternberg, R. J. (2015). Successful intelligence: A model for testing intelligence beyond IQ tests. *European Journal of Education and Psychology, 8,* 76-84.

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

- Umaña-Taylor, A. J. (2016). A Post-Racial Society in Which Ethnic-Racial Discrimination Still Exists and Has Significant Consequences for Youths' Adjustment. *Current Directions in Psychological Science, 25*, 111–118.

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

- Weiner, J. (1994). *Physical Abilities Test Follow-up Validation Study*. Sacramento, CA: The California Commission on Peace Officer Standards and Training State of California.

References

- Wenger, J. W. (2010). *Expanding the Recruiting Market: Noncognitive Testing; CRM D0023223.A2/Final*. Marine Corps Manpower Team, Resource Analysis Division.

References

- White, L. A., Rumsey, M. G.; Mullins, H. M., Nye, C. D. & LaPort, K. A. (2014). Toward a new attrition screening paradigm: Latest Army advances. *Military Psychology, 26,* 138-152.

References

- Wiesen, J. P. (2016). Tools to increase diversity and validity in hiring police officers. *The Personnel Testing Council of Metropolitan Washington. Newsletter, XII,* 4-11.

References

- Wiesen, J. P. (2017). Tools to increase diversity and validity in hiring police officers - Part II. *The Personnel Testing Council of Metropolitan Washington. Newsletter, XII*, 6-15.

References

- Wiesen, J. P. (2017). Tools to increase diversity and validity in hiring police officers –Part III. *The Personnel Testing Council of Metropolitan Washington. Newsletter, XIII*, 6-17.

References

- Wiesen, J. P. & Aguinis, H. (2010). *New Methods for Reducing Adverse Impact and Preserving Validity.* Symposium presentation at the 25th Annual Convention Society for Industrial and Organizational Psychology, Atlanta, GA.