Content Validity Court Decision: Tatum, 2022; An Insider's Analysis and Insights

Joel P. Wiesen, Ph.D.

Contact: **jw@jpwphd.com**

39th Annual SIOP Conference
Chicago, IL, 4/18/2024

Who Am I?

- Independent consultant
- 45 years as expert for Defense and Plaintiff
- 17 years as expert related to *Tatum* case
- Worked for the MA civil service agency
- A scientist-practitioner
- Ph.D. experimental psych, minor in psychometrics/statistics

---

What is the *Tatum* Case?

- Challenged 8 sergeant promotion exams
- 2 exams for Boston
- 6 annual "statewide exams"
  - Same exam for 100+ PD in Massachusetts
- Plaintiffs prevailed

Goals of This Tutorial

- Provide an overview of the *Tatum* case
- Discuss main findings and disputes
- Identify testing weaknesses seen in *Tatum*
- Propose ways to address these weaknesses
- Propose new approach to police promotion

Organization of Presentation

A. Tatum complaint
B. Background to case
C. Major topics of testimony
D. Court decision
E. Where experts differed and court decision
F. Issues and proposed solutions
G. New approach to improve such exams

A. *Tatum* Complaint

- Complaint in 2009, case decided 2022&23
- Plaintiffs were Black and Hispanic candidates
- 8 promotional exams for sergeant in MA
- Impact discrimination
- Class action: Black and Hispanic test takers

B. Background Topics

- Role of a municipal police sergeant
- Previous MA impact discrimination cases
- Related MA police promotion impact cases
- Overview of MA police promotion exams
  - Exam components
  - Passing points
  - Number of exams challenged

Role of a Police Sergeant

- Sergeants supervise officers (about 7)
- Sergeants spend much time in the field
  - Go to the most serious incidents
- No sergeant at the Geo. Floyd incident
  - Was a minor crime of passing a fake $20 bill
- Officers ask sgts. if uncertain (law, SOPs)
- Sgts must answer quickly (instantaneously)

MA Impact Discrimination Cases

- History of MA impact discrimination cases
  - Police officer, 1971
  - Firefighter, 1974 and 2006
  - Police sergeant, Boston, 1978 (consent decree)
  - Remedial hiring preferences for decades

Related Police Promotion Cases

- Two closely related impact cases in MA
- *Lopez v Lawrence*, 2014 (trial in 2009)
  - Same claim, but in federal court
  - Court ruled exams **were valid**
- *Smith v Boston*, 2015
  - Same claim in federal court but for lieut. exam
  - Court ruled 2 Boston exams were **invalid**

Lopez 2014

- Court ruled exam was valid despite AI
  - Why?
- Expert for Defense:
  - A) MC test alone was not valid
  - B) With E&E the exam was "minimally valid"
  - C) No guarantee alternatives would have less AI
- Court ruled A and B, did not get to C

Why 2 Trials on Same Exams?

- Different defendants
  - Several cities in *Lopez*
  - MA Human Resources Division in *Tatum*
- MA & federal courts use the same criteria
  - Uniform Guidelines, etc.
- *Lopez* complaint originally included HRD
  - HRD dismissed: state not Title VII employer

Smith 2015

- Court: exam measured narrow slice of job
- "As discussed above, the test outlines reflect that the 2008 exam was written to test knowledge; the outlines reveal that only two **abilities** appeared on the 2008 exam."

Smith 2015

"… the Court has found that the E&E component had very little bearing on the final score, the Court cannot find that the City met its burden … too many skills and abilities were missing from the 2008 test outline."

Overview of MA Exams

- Civil service exams in MA are centralized
- 100+ statewide police departments (PDs)
- PDs participate biannually, typically
- Most PDs make few, if any, promotions
- Most PDs have few, if any, B or H takers
- Promotions made in order of rank on exam
- Exams developed in-house by MA HRD

Exam Components

- Police promotion exams had 2 components
- 80%: MC exam based on textbooks and law
- 20%: Ed & Exp. (E&E) points for years
- Both graded on a 100-point scale
- One passing point, typically 70%

Multiple Choice (MC)

- M/C questions based on written sources
- Textbooks for statewide exams, incl. law
- SOPs only used for Boston exams
- Many **verbatim** quotes from sources
- Many **definition** questions
- No problem solving

Education and Experience (E&E)

- Entrance requirement: 3* years as an officer
- Minimum possible E&E score is 70
  – Screened for entrance requirement before exam

* 1 year if not enough candidates

Exam Passing Points

- Passing point 70% for 7 of 8 exams
  - One overall passing point for each exam
- Passing point not correlated with difficulty
  - Most difficult exam had passing point of 70%
- Passing point set, in part, to reduce AI
  - Passing point of 65% in 2008

Number of Exams Challenged

- 8 exams held from 2005 to 2012
- 6 statewide exams challenged (not 09 &11)
- 2 Boston exams challenged
- 57 PDs partook in at least 1 statewide exam
- 27 PDs partook in more than 1 exam
- 96 PD-exam year combinations in all

C. Major Topics of Testimony

- Adverse impact (AI)
- AI for promotion, pass rate, means
- How or whether to do AI analyses across PDs or exams (years)
- Limitations of MC job knowledge (JK) items
- Validity of Education & Experience (E&E)

More Major Topics of Testimony

- Reliance on SMEs
- Passing points
- Ranking vs. banding
- Expected adverse impact of alternatives

D. Court Decision

- Insightful evaluations by a seasoned judge
- Some topics in the decision
  - Conclusions
  - Adverse impact
  - Validity
  - Passing points
  - Alternatives

Decision: Conclusion

"**Overwhelmingly persuasive evidence** proves that HRD interfered with the class members' rights to consideration for promotion to police sergeant without regard to race or national origin."
[emphasis added]

Decision: Conclusion

"…a discriminatory system that has injured qualified candidates and **deprived the public of the benefits of having the best-qualified police sergeants.**"
[emphasis added]

Decision: Conclusion

"In all these actions, **HRD knew what it was doing."**
[emphasis added]

Decision: Adverse Impact

"…regularly administered written exams, **knowing** that its testing format had an **unnecessary, plain and obvious adverse impact** upon Blacks and Hispanics…"
[emphasis added]

Decision: Validity

"Instead of improving its assessment format, HRD promulgated lists to provide **a thin veneer of apparent justification** for a discriminatory process."
[emphasis added]

Decision: Passing Points

"HRD … did not rely on any accepted scientific criteria for establishing the passing score for its exams."

Decision: Alternatives

"HRD **knew of clearly superior assessment methods,** but **continued to use the same, unnecessarily discriminatory format** anyway."

"HRD **failed to implement some very simple ways to reduce adverse impact** upon Black and Hispanic candidates."
[emphasis added]

Summary of Tatum Decision

- Used job knowledge tests for many years
- JK tests consistently had adverse impact
- Intent discrimination based on past impact
- JK tests measured rote memorization
- JK tests did not measure important KSAPs
- JK tests invalid, especially for ranking
- Did not use alternatives with less AI

Decision: Rationale

- "The court finds … by a **preponderance of the evidence** it finds **credible**…"
  [emphasis added]
- Will discuss the evidence next

E. Contested Topics Format

- Defense's position
- Plaintiffs' position
- Court decision

11 Contested Topics

- Aggregation
- AI: Means
- AI: Promotions
- Study Sources
- SME input on items
- Job analysis

11 Contested Topics (continued)

- Item content
- Education & Experience methodology
- Ranking vs. banding
- Passing Points
- Alternatives

Aggregation: Defense

- Omit PDs with no promotions from statistical analyses of means
- Do not aggregate across years because unknown number of retakers violates the assumption of independence
- Do not aggregate across PDs because mean scores may differ by PD

Aggregation: Plaintiffs

- Include PDs with no promotions to increase power when comparing means
- Defense repeatedly refused to provide candidate IDs making it impossible to do analyses omitting retakers
  - Candidate IDs never produced in usable form
- Aggregate across PDs to see the big picture

Aggregation: Court

"The court rejects [Defense's] position that it should disregard entirely any results that are not based upon a single test in a single department, after excluding all departments that had no diversity or made no promotions. [That] approach is **biased in favor of finding no difference** in treatment of White, Black and Hispanic candidates."

Adverse Impact **Means**: Defense

- Requires a significant 3 group ANOVA (B, H, W) before testing Minority-White
- Bonferroni approach used for B-W, H-W, M-W follow up tests
  - Found no AI for most PDs

Adverse Impact **Means**: Plaintiffs

- Mean scores germane to demonstrating AI
- The test statistics show a pattern
  – Exam scores drive P/F and promotion rates
- Power greater for means than promotions
- No reason to use a Bonferroni approach
- Many statistically significant differences between M-W means when tested correctly

Adverse Impact **Means**: Court

"Dr. Wiesen found statistically significant mean score differences favoring Whites over minorities in ... [11 PD-year combinations]."

Adverse Impact **Ratio**: Defense

- No promotion AI for most of the 8 exams
- No statistical significance in small munis
- "Shift of one" avoids 80% for small munis
  – **Used shift of one 30 times**
- Collapsing across munis illogical
- Collapsing across years statistically wrong
  – Retakers violate statistical assumption

Adverse Impact **Ratio**: Defense

- **Simpson's Paradox** could be operating
  - Threat of misinterpretation of some data patterns

Simpson's Paradox: Within Muni

Muni	Group	Outcome		Promo Rate
		Not Promo	Promoted	
Muni A	Minority	1	1	0.5
	White	10	10	0.5
Muni B	Minority	18	2	0.10
	White	45	5	0.10

Simpson's Paradox: Across Munis

Group	Outcome		Total	Promo Rate
	Not Promo	Promoted		
Minority	19	3	22	0.14
White	55	15	70	0.21
AI Ratio				**0.64**

Adverse Impact Ratio: Plaintiffs

- Significant promotion AI in largest muni
- **Zero power** for munis with only 1 M taker
- **Zero power** for munis w/ only 1 promotion
- Collapsing across munis logical for promos
  - Mantel-Haenszel test appropriate
- Look at pattern of data even if AI not signif.
  - Means drive P/F and promotion ratios

Adverse Impact Ratio: Plaintiffs

- Shift of one intended to be used
  "**in the absence of other evidence.**"
  - Uniform Guidelines Q&As, March 2, 1979
- "Other evidence" includes:
  - Pattern of severe AI for many decades
  - Mean score differences are statistically signif.
- Using "shift of 1" for 30 munis was wrong

Adverse Impact Ratio: Plaintiffs

- **Simpson's Paradox not seen in case data**
  - Pattern of ratios the same overall and in PDs
  - No actual misinterpretation of ratios
  - Defense raised issue of Simpson's Paradox to try to mislead court

Adverse Impact Ratio: Plaintiffs

- Collapse across years was thwarted by Defense by not providing candidate IDs

Adverse Impact Ratio: Court

- "P-values greater than .05 but well below 1.0 … [do] not mean that the data lack all meaning, or that a court should exclude the data from consideration as part of a larger body of evidence."

Adverse Impact: Court

- "The massive amount of evidence proving the known and unjustified disparate impact of HRD's format leaves no doubt in this court's mind…"
- "…plain and obvious adverse impact upon Blacks and Hispanics"

Study Sources: Defense

- Panel of chiefs approved reading list

Study Sources: Plaintiffs

- College textbooks with academic focus
- Much irrelevant content
- No SOPs for any PD except Boston
  - 100+ PDs have no test of knowledge of SOPs
  - Knowledge of SOPs essential for a sergeant

Study Sources: Court

"Shortening the reading list would also reduce the cognitive load. It was not necessary, for instance, to include a **600-page** Police Administration textbook on the 2008 reading list, when the exam included only 2 questions from that book on the statewide exam and **1 question** on the Boston exam."

SME Input on Items: Defense

- 2 or 3 SMEs reviewed each test question
- Usually at rank of chief

SME Input on Items: Plaintiffs

- SMEs asked to identify easy items
- Easy items were prioritized for inclusion
  – 70 easy items used on one 80 item exam
- Exams did not reasonably cover the KSAPs

SME Input on Items: Court

"HRD … appears to have used only two SMEs … to review reading lists and examination questions. With so few SMEs and given the deficiencies identified below, the court gives only modest weight to the SME process in assessing the validity of the exams for statewide application or use in Boston."

Job Analysis: Defense

- Test outline was based on job analysis
- MC questions covered many essential KSAPs including law
- E&E measured other required KSAPs

Job Analysis: Plaintiffs

- Major basis for the exams was 1991 study
  – Job analysis from 14 to 21 years old
- Newer, 2000 Boston job analysis flawed
  – **Apparently manipulated**

Job Analysis: Plaintiffs

- **1991 JA study apparently manipulated**
- 11 SMEs chosen for diverse assignments
- "How often do **you perform** this task?"
- SMES made some 26,000 ratings of aspects of 218 job tasks and 165 KSAPs
- The 11 SMES agreed perfectly
  – Despite different job assignments

Job Analysis: Plaintiffs

- Illogical ratings of tasks
  - Qualify/practice with weapons daily
  - Talks with leaders of demonstrations daily

Job Analysis: Plaintiffs

- Majority of important KSAPs not tested
- M/C did not measure situational judgment, interpersonal skills, ability to plan, reach logical conclusions, etc.
- PD SOPs not measured (outside Boston)

Job Analysis: Court

"…the SMEs ranked the importance of various KSAs as part of that job analysis. Their rankings are implausible. The 11 SMEs gave identical rankings to all of the approximately 1,100 ratings…"

Job Analysis: Court

"…[the] job analysis also claimed that police sergeants perform certain tasks every day, but that could not possibly be true …

– Practice in operation of firearms/weapons
– Set up command post
– Directs major incidents
– Inspect licensed premises"

Job Analysis: Court

"The court rejects the inference that these test development processes, by themselves, "ensured that those exams were job related"''

Items: Defense

• M/C questions measured situational judgment, observation skills, interpersonal skills, ability to plan, reach logical conclusions based on information at hand
• M/C questions tested reading sources
• Items reviewed by SMEs (chiefs)

Items: Plaintiffs

- Items tested **rote memory**, not application
- SME item review was flawed

Items: Court

"…many questions are **definitional** in that the answers turn upon the meaning of a particular word."

"Those questions have **low fidelity**, because a sergeant's job does not generally involve using academic jargon or other definitions of concepts in the assigned reading."

Items: Court

"According to CP, the most critical determinant of future success as a community policing Officer is:

A. Superior communication skills.

B. Empathy. (key)

C. Autonomy.

D. Analytical ability."

Items: Court

- Defense said this item measured empathy
- The court critiqued this item as only measuring knowledge about empathy but not the ability to be empathetic or foster empathy in subordinates

---

"According to PDRCIP, regarding the requirements for a search to be legally valid, it would be correct to state that:

A. Neither the search warrant nor the accompanying affidavit are required to be brought to the scene.

B. The affidavit must be brought to the scene and presented on demand, to the owner or occupant of the premises, but it is not required to bring a copy of the search warrant."

"C. The search warrant must be brought to the scene and presented on demand, to the owner or occupant of the premises, but it is not required to bring a copy of the affidavit.

D. Both the search warrant and the affidavit must be brought to the scene but the police are not required to present either document to the owner or occupant of the premises, even upon demand."

Items: Court

"[The SME] Chief ___ agreed that it would "make more sense to ask the sergeant candidate what's good practice to bring to [… the scene] as opposed to what's legal, because it might be that good practice is more than what's required by the law.""

Items: Court

"…information ... unrelated to the sergeant's job ... the maximum length of prison sentences allowed by law for certain offenses"

Items: Court

"…inability of a written multiple-choice exam to predict good job performance as a sergeant."

"…questions on the exam **largely test for rote memorization** of facts and passages…"

Items: Court

"…the exams did not test many important job qualifications."

"…not measure ability to **apply knowledge practically** and to **exercise judgment** on that topic **in specific situations**"

Items: Court

"**Chief among the essential skills** … are:
Leadership skills,
Supervision skills,
Decision-making and problem-solving,
Interpersonal skills,
Communication skills, and
Integrity."

Items: Court

"[HRD] failed to test meaningfully the KSAs required for good performance as a police sergeant."

"There is no credible evidence that [the exams] evaluated which information police sergeants must memorize in order to perform their job."

Items: Court

"Most of the questions on the exams at issue in this case tested topics that were important to the job of sergeant. That does not mean that HRD's format was reasonably job related. It was not."

Education & Experience: Defense

- Point-year system of E&E
- E&E claimed to measure many KSAPs:
  - Perceiving/reacting to the needs of others
  - Ability to write, prepare reports
  - Ability to be confidential
  - Ability to follow policies and procedures
  - Ability to interpret policy

Education & Experience: Plaintiffs

- No validity study for E&E
- E&E not designed to measure KSAPs
- E&E not linked to specific KSAPs or duties
- Years of exp. not linked to KSAPs/duties
- Amount, not quality, of experience credited
- No credit for experience outside a PD
  – Many logical jobs omitted from E&E

Education & Experience: Plaintiffs

- No valid basis for 20% weight for E&E
- Actual weight of E&E was small
  - Mathematically: 6 points out of 100
- Correlation between MC and exam score high
  - .97 and .98 for the two Boston exams

Education & Experience: Court

"No empirical support or credible professional study justified the 20% weighting."

Education & Experience: Court

"…the effective weight of E&E component is substantially lower than 20% because of the way HRD scores E&E."

"…the final scores on HRD's exams correlated in a perfect linear relationship with the score on the multiple choice tests…"

Education & Experience: Court

"…limited E&E component"

"…essentially the same today as it was 50 years ago"

"HRD did not in fact capture these skills in the E&E component."

Education & Experience: Court

"…no credit for ... community policing or involvement in the communities served"

"…no credible support for the notion that a bachelor's degree was the equivalent of six years job experience."

Education & Experience : Court

"…[There were] serious flaws in identifying (1) which KSAs are testable on a multiple-choice exam and

(2) which KSAs are measured in HRD's education and experience component."

Ranking: Defense

- Ranking required by state law
- Exams measure **some essential** KSAPs
- Exams measure much of job
  – 40% of the KSAs could be tested by written test
- How much candidates know is important
- Sergeants must know law (25% of MS test)

Ranking: Plaintiffs

- Ranks are imprecise predictors of job perf.
  – Did not measure enough of job for ranking
  – Other KSAPs drive job performance also
- Only 22% of KSAPs measured by exams
- Some law questions impractical
- Definition and memory questions weak
- State itself advocates and uses banding

Ranking: Court

"HRD failed to test many important KSAs … and failed to ask questions that focused upon measuring job-related knowledge, its format did not rank candidates for promotional purposes on a basis that was substantially job related."

Ranking: Court

"No credible study showed that single-point differences in scores reflected any significant difference in job qualifications."
"HRD itself ... proposed 'banding' ... has already conceded that its multiple-choice examinations were not sufficiently valid as rank order devices, even though they now claim just the opposite."

Pass Point: Defense

- Passing points not important with low selection ratio (i.e., in large PD)
  - Never reach people near the passing point

Pass Point: Plaintiffs

- No attempt to link passing point to job
  - Not satisfy state law: identify qualified people
- Passing does not indicate competence
  - Many easy items
- Many Incumbent Sgts. fail the test for Sgt.
- Passing point not job related

Pass Point: Court

"[HRD]…did not rely on any accepted scientific criteria for establishing the passing score for its exams."

"…incumbent sergeants who take HRD promotional exams for lieutenant often do not perform well on the sergeants' portion."

Alternatives: Defense

- Alternatives do not guarantee less AI
- Alternatives can increase AI

Alternatives: Plaintiffs

- Expert's firm pushes alternatives
  – In general, alternatives improve validity and reduce AI

Alternatives: Court

"As [Defense's expert] acknowledged at trial, performance review systems 'can be useful and they do tend to reduce adverse impact.' His own company … has recommended use of such systems."

Alternatives: Court

"HRD knew of clearly superior assessment methods, but continued to use the same, unnecessarily discriminatory format anyway."

Transition

jpwphd.com/siop2024

F. Issues and Proposals

- Some major issues/flaws in testing
- Some little discussed
- Based on *Tatum* and other exams reviewed
- Approach of the rest of this presentation:
  – Describe issue
  – Offer a proposed solution

Topics of Issues

- Items
- SMEs
- Job analysis
- Test outline

Issue: Allowing Item Appeals

- Some civil service rules allow item appeals
- Appeals typically heard by a lay body
- Item upheld if it closely reflects the source
- So, items quote sources
- Items measure recall of wording in source
- Little measurement of practical application

Proposal: Allowing Item Appeals

- New law/rule to grant authority to SMEs
  - e.g.: Post-test agreement of 3 SMEs presumed to be adequate support for an item (SMEs who did not write the item)
  - Involve police academy, municipal attorney
- Concern: Candidates will say they cannot study for SME questions

Issue: Replicate Legal Cases

- Law questions are common on police exams
- Questions typically completely replicate **all** the facts of an actual court case
- This avoids appeals
- Does not replicate job duties of sergeant
- Officers must respond to incidents that do not fully replicate past court cases.

Proposal: Replicate Legal Cases

- Use items that do not fully replicate past cases
- Require judgment in applying precedents
- Involve local attorney in item development
  - City attorney, etc.

Issue: Definition Items

- Easy to write & defend definition questions
- Knowing a definition does not mean person can use the concept

Proposal: Definition Items

- Use only a small proportion of definition items for any given KSAP
- Use definition item only if the definition is important to know in order to do the job

Issue: Items on Procedures

- Easy to ask name or order of procedural step
- Knowing the name or correct order of steps does not mean person can execute the steps
- Q: What should be done in the analysis step of problem solving?
  - key: Collect information from a variety of public and private sources
  - Does not test if able to collect information

Proposal: Items on Procedures

- Test ability to implement the steps
- Test the name or sequential order of steps only if these are important (e.g., step is likely to be done out of sequence)
- Use only a small proportion of step name or sequence items for any given KSAP

Issue: Academic Items

- Question on desired leadership style
  - Key: balances concern for people and task
- A correct answer does not mean the person can do either well

Proposal: Academic Items

- Use academic items only if application clear
- Items on important, applied topics
- More situational questions
  - Video stimuli
  - Constructed responses
- May require item writers to have more skills
  - Work with actors and video content creators

Issue: Limited Item Review

- Tatum item review questions 2008 Boston:
- Suitability for rank?
  – No definition of suitability
- Estimated difficulty "for the persons taking the examination"
- Estimated readability

Proposal: Limited Item Review

- Gather better information from the SMEs
- Improve the item rating form
- Clarify the review topics and rating levels
  – Is this K important to do the job?
  – Is this the best way to measure this knowledge?
- Talk with SMEs about each item.
  – How is this knowledge used on the job?

Issue: Job Analysis Accuracy

- Task and KSAP inventories with implausible results
    Tasks not done daily; KSAPs omitted
- Major disagreement among SMEs
- Illogical ratings:
- Tasks of budgeting; read, interpret tables/graphs: but no math ability required

Proposal: Job Analysis Accuracy

- Do not blindly rely on SME ratings
- Probe discrepant and suspect ratings
- Review the JA results for plausibility
- Conduct reviews of JA findings with SMEs
- Gather ratings on KSAPs from
  - Supervisors
  - Training academy staff

Issue: KSAPs Not Well Defined

(A) K of principles of management
versus
(B) POSDCORB areas listed separately
- If (A), are all SMEs rating the same area?
- If (A), how much emphasis on each facet?
- Lack of clarity affects job analysis
- Lack of clarity affects test outline

Issue: KSAPs Not Well Defined

- Consider these 3 K statements from a recent job analysis for police lieut. and captain:
- Principles of police administration
- Supervision, management, and leadership principles
- Community-policing and problem-solving principles

Issue: KSAPs Not Well Defined

- What do these Ks cover, what is being rated
- Which community policing practices?
- What if some principles are very important and others not?
- Asking SMEs to interpret job analysis results may not work

Proposal: KSAPs Not Well Defined

- Ask if every SME will agree to scope
- Ask if the KSAP can be broken down into components that are not highly correlated
- Use operational definitions of KSAPs

Issue: Many Tasks/KSAPs

- Often there are many tasks and KSAPs
- Grouping into task categories loses detail
- Grouping into competencies loses detail

Proposal: Many Tasks/KSAPs

- Use tasks and KSAPs when writing items
- Do not rely on task/KSAP groupings

Issue: KSAP Weight

- Rate KSAPs with **ordinal** scale
- Average importance informs test plan
- But no standard size of a KSAP
- Some KSAPs encompass much information
- Some sources are many pages
- Generating items easier for some sources

Proposal: KSAP Weight

- Rate KSAPs with **ratio** scale
- How much of successful job performance depends on this KSAP?
- Allot 100 or 1,000 points among the KSAPs
  - Use Excel to ease math burden
- Frank Landy used this approach in his job analysis of police officer in Massachusetts

Issue: Critical Incident Usage

- List of tasks (briefly stated) is inadequate
- Short "ride-a-longs" are inadequate
- Often critical incidents are not collected

Proposal: Critical Incident Usage

- Collect critical incidents from incumbents
- Collect critical incidents from supervisors
- Goal: Many hundreds of incidents
- Incidents provide grist for item writing

Issue: Outline Based on Tasks

- Typically test outlines are based on KSAPs
- KSAPs are a step removed from the job
- Sources for KSAPs can be quite academic
- How to tell if a test is representative of job?
  - KSAPs are a step removed from tasks

Proposal: Outline Based on Tasks

- Develop test outline based on tasks
- Can have 2-way outline
  - sources and tasks
  - KSAPs (or KSAP groupings) and tasks
- Easier to show test is representative of job

Issue: Weighting Items

- Score on exam is typically # correct
- Easy and hard items have same weight
- Some areas tested with few items
  - No reliability measure of such areas

Proposal: Weighting Items

- Enough items to reliably measure a KSAP
- Weight that measure by KSAP importance

Issue: Setting Passing Point

- Angoff rating is compensatory
  - What if easy questions are in an essential area?
- Illusive "minimally qualified incumbent"
  - This is indirect rating of passing point

Wiesen (2024) SIOP Conference   130

Proposal: Setting Passing Point

- Consider alternatives to Angoff question
- Rate questions as an entity
- How many of these items would have to be answered correctly to indicate a person can do the job?
- Do this by test area
- Could also ask about missing KSAPs

Wiesen (2024) SIOP Conference   131

Issue: Single Pass Point

- Items weighted equally
- Grading compensatory
  - Can pass w zero K of some area
- Can pass exam with zero on a KSAP
- In small muni., maybe only 1 passes exam
  - That person will be promoted

Wiesen (2024) SIOP Conference   132

Proposal: Single Pass Point

- Identify essential KSAPs
- Set passing point for each essential KSAP

Issue: Test Outline Secrecy

- Some testing groups do not reveal outlines
- Often voluminous sources
- Candidates do not know what to study most
- Identifying material to study not job related

Proposal: Test Outline Secrecy

- Tell candidates the # items per source
  - Allows candidates to apportion study time
- Give guidance on what will **not** be tested
  - History older than 10 years
  - Chapters x, y, z

Issue: Professional Secrecy

- There are no compendia of:
  - Test outlines
  - BARS scales
  - Practical exercises
  - M/C job knowledge questions

Proposal: Professional Secrecy

- Perhaps SIOP could publish compendia

Issue: BARS Reliability

- Within board reliability: 0.97, 0.91
- Between board reliability: **0.44**
  - .97 x .91 = **.88** (n>100)
- Within board reliability: 0.97, 0.94
- Between board reliability**: 0.66**
  - .97 x .94 = **.91** (n>100)

(Note: data not from MA)

Proposal: BARS Reliability

- Use multiple rating board more often
- Research into reasons for disagreement

G. New Approach to Exams

- For police promotional exams
- Several goals of the new approach
- Better job performance of sergeants
- Reduced adverse impact
- Better acceptance of promotion process
  - Candidates
  - Management

Issue: New Approach Needed

- Candidates prepare themselves for promo.
- Hard to learn supervision, management, leadership, strategy, tactics, etc. from books
- Exam grades show candidates lack KSAPs
- Many high paying occupations have training programs
  - $200,000+ average gross pay, Boston Sergeant

Issue: New Approach Needed

- Currently, many/most exam scores are low
  - A low-scoring candidate may be promoted
- Candidates weak in essential KSAPs
- No training for newly promoted sergeants

Proposal: New Approach Needed

- Establish a thorough training program for promotional candidates
  - Strategy and tactics for incidents
  - Planning and resource allocation
  - Interpersonal aspects of policing
- This is a major undertaking
  - Fund course development
  - Fund training time for candidates

In Closing

- We can improve police testing practices
  - Better content valid knowledge tests
- Both candidates and management benefit
- Avoid or win more testing court cases
- **Improve job performance of sergeants**

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

Contact Information

- Session URL: **http://jpwphd.com/siop2024**
- Email: **jw@jpwphd.com**
- Telephone: **(617) 244-8859 (land/no text)**
- **Email and telephone calls welcome!**
- **Q&A's**

Q&A

Main Cases Cited

- *Lopez v. Lawrence*
- Lopez v. City of Lawrence, Civil Action No. 07-11693-GAO (D. Mass. Sep. 5, 2014)
- https://casetext.com/case/lopez-v-city-of-lawrence-1?q=07-11693-GAO&sort=relevance&p=1&type=case

Wiesen (2024) SIOP Conference   148

Main Cases Cited

- *Smith v. Boston*
- Smith v. City of Boston, 144 F. Supp. 3d 177 (D. Mass. 2015)
- https://casetext.com/case/smith-v-city-of-bos-1?q=12-10291-WGY&sort=relevance&p=1&type=case

Wiesen (2024) SIOP Conference   149

Main Cases Cited

- *Tatum v. Massachusetts*
- Tatum v. Mass., C.A. 0984CV00576 (Sup. Court 2022)
- https://www.mass.gov/doc/tatum-et-al-v-human-resources-division-related-superior-court-decision-102722/download

Wiesen (2024) SIOP Conference   150

Presentation Citation

Wiesen, J.P. (2024). Content Validity Court Decision: Tatum, 2022; An Insider's Analysis and Insights [Master Tutorial]. *Society for Industrial and Organizational Psychology Annual Conference*, Chicago, IL, United States.