

NYC's Firefighter Exams in Federal Court, as Seen by One of the Plaintiffs' Experts

Joel P. Wiesen, Ph.D.
wiesen@appliedpersonnelresearch.com
PTC/MW Luncheon, 5/21/14
Arlington, VA

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 1

Why Did You Invite Me?

- Insider's view of a prominent case
 - Why decided on summary judgment in 2009
- What happened since 2009
 - **Outsider's** information on most recent exam
- What we can learn from the case
 - Not as much as we would like!
 - Read Judge Garaufis' Memorandum and Order

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 2

Today's Presentation

- I Overview of Case
- II Adverse Impact
- III Validity
- IV Alternative Selection Procedures
- V Court Decision
- VI What Happened Next (2012 Exam)
- VII Gleanings

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 3

I. Overview of Case

- Snapshot of case
- "Typical" Title VII Impact Case
 - Except there are two Plaintiffs
- Job is Firefighter
- Decided on Summary Judgment
 - Adverse impact was present
 - Insufficient evidence of business necessity

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 4

Snapshot of Case

- Challenge to NYC Firefighter Entry Exam
 - minimum qualifications
 - written test
 - Physical Performance Test (PPT)
- Case brought by DOJ: Blacks and Hispanics
- Vulcan Society intervened for Blacks
 - challenged more parts of exam

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 5

"Typical" Title VII Impact Case

- Adverse Impact
- Validity
- Alternative Selection Procedures

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 6

Job of Firefighter

- Very hands-on
- Teamwork important on fireground
- Supportive environment in station
- Supportive environment in academy
- Verbal communication important
- College is not typical nationwide

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 7

Four Challenged Practices

- P/F use of the two exams
- Ranking based on the two exams

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 8

Decided on Summary Judgement

- Unusual for Title VII case
- No “triable” issues
- Court order and memorandum of 2009
 - Very clear statement of facts and reasoning
- We will review the court’s reasoning

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 9

II. Adverse Impact

- Minimum Qualifications
- Adverse Impact at nominal passing point
- Adverse Impact at effective passing point
- Adverse Impact within top PPT score

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 10

Minimum Qualifications (MQs)

- 30 college credits (any courses)
- Driver’s License
- Certified First Responder Certification with Defibrillation
- In short, there is no validation of these MQs

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 11

Validity of 30 College Credits

- City employee said:
“... the education requirement...was something that Commissioner [name redacted] was extremely hepped up about. He insisted that an education requirement be added. What Commissioner [name] wanted, Commissioner [name] got.”

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 12

Driver's License

- Adverse Impact
- Validity
 - Driving task dropped out in the job analysis.
 - Most Firefighters do not drive on the job

Certified First Responder Certification with Defibrillation

- Minimum Qualification
 - 1999 exam: required to take exam
 - 2002 exam: required by the end of the academy
- Two key topics:
 - Adverse Impact
 - Validity

III. Validity

- City's approach to validation
- City's validation reports
- Major testing flaws

Overview of Validation Approach

- Little change to past test development
- Fleishman abilities
- SME job analysis ratings (N=192)
- SME linkage ratings (N=12)

City's Validation Reports

- One test development report
 - Mainly job analysis and test plan
- First exam: 9 page report plus 9 appendices
- Second exam: **No test development report**
“probably can use old job analysis”

Plaintiffs' Experts' Reports

- Adverse Impact (Siskin, Wiesen)
- Job analysis (Goldstein, I., Wiesen)
- Validity (Hough/Jones, Wiesen)

Defense's Experts' Report

- Adverse Impact and validity (Bobko, Schemmer)
 - They did not develop the 1999 or 2002 tests.

Major Testing Flaws

- Test design/validation flaws
- Test development flaws
- Test question flaws

Test Design Flaws

- Aspects of the Physical Performance Test
 - Will not discuss today
- Used Fleishman's 18 cognitive abilities
- Limited KSAPs to Fleishman abilities
 - SMEs could not add to ability list

Fleishman Abilities Not Understood

- "Linking Panel" SMEs did ability-task linkages and rated abilities
 - 3/4 of these **SMEs did not understand** some (many?, most?) of the Fleishman abilities
- Examiners who wrote questions also **did not fully understand** the Fleishman abilities

SME Deposition: Fleishman Abilities

- SME example of Perceptual Speed: Remembering the location of objects in a smoky room
- But this is an example of Memorization

SME Deposition: Fleishman Abilities

- SME example of Timesharing:
A situation where one firefighter wanted to go home early and another firefighter was willing to come in early to cover for him.
- But this is not an example of Timesharing.
- This SME said he did not find any of the ability area definitions confusing.

SME Deposition: Fleishman Abilities

- SME deposition:
- Q. With respect to the part of understanding what the person was writing on the board during the drill, would you say that's written expression or written comprehension?
- A. I'm not sure.

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 25

SME Deposition: Fleishman Abilities

- SME deposition:
- After having the definition of Deductive Reasoning read to her and on being asked how important it is to the job, one SME said:
"Again, I don't understand the definition. I'm sorry."

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 26

Lack of Care in SME Ratings

- One SME rated Written Expression as "Important" for 4 tasks that do not involve writing:
 - Climbing and portable ladder activities
 - Building entry
 - Search
 - Extrication

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 27

SMEs Disagree on Ability Ratings

- Disagreement on Written Comprehension for Size Up
- 3 SMEs rated Written Comprehension as 0 or 1
- 4 SMEs rated it 3 (the highest rating)

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 28

Test Development Flaws

- Job analysis
- Test questions

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 29

Job Analysis

- SME understanding of Fleishman Abilities
- Lack of internal consistency in SME ratings
- MQs and PPT not considered in JA

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 30

Representativeness

- Q. In your opinion, of all the abilities that are required by or that are important to the entry level firefighter job, what percentage do those nine abilities represent?
- A. [Dr. Bobko] I don't know.

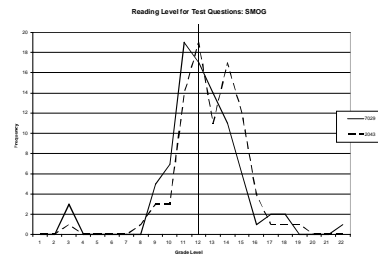
Test Question Flaws

- Link to test areas
- Reading level
- Factor structure

Link to Test Areas

- Court questioned this, as discussed below

High Reading Level



Hard Words

accordance	constitute	fluctuating	partitioned
adjoining	debris	imminent	procurement
adjoins	emit	instituted	profusely
allegation	enables	inverted	solely*
amended	ensure	obligation	stabilize
ascend		paraphernalia	subsequently

* 27 times

Measuring Unintended Abilities

- 5 or 7 items measure math (in the 2 tests)
- Math is not on test outline

Wording and Adverse Impact

- Longer questions are harder
 - Greater effect for blacks
- Questions with jargon are harder
 - Greater effect for blacks
- More blacks than whites left last questions blank

Factor Structure

- Court questioned this, as discussed below

Some Topics Not Covered Today

- PPT
 - Design
 - Implementation

Summary of Plaintiffs' Expert Reports

- Job analysis
- Minimum qualifications
- Test validity (including test construction)
- Combining scores
- Cut scores
- Ranking
- Alternatives

Job Analysis (JA)

- JA used Fleishman cognitive abilities
- JA ignored non-cognitive mental abilities
- JA ignored physical
- JA ignored MQs

Test Validity

- City's experts did not prove validity of:
 - test content
 - cut point
 - ranking

Test Construction

- Item review panel (4 FF, 1 Lt) not given criteria for item review and selection
- Adverse impact of 1999 exam did not result in changes in 2002 exam

Cut Scores

- No validation support provided by City
- Arbitrary
- Different passing points for the two exams

Ranking

- No validation support provided by City

Alternative Selection Procedures

- Dr. Schemmer said tests exist for teamwork, responsibility, getting along with others, etc.

Opinions of Experts for Defense

- Defense's expert report and depositions
- Minimum qualifications
- Validity
- Cut-off
- Ranking

Defense Experts' Reports

- Length: 33 pages
- Content
 - 20 pages on adverse impact
 - 3 pages on validity

Defense on Adverse Impact

- Attacked nature of statistical testing:
- Improper to use parity as the null hypothesis
 - “differences in aptitude ...[are] a matter of business necessity”*
- Large Ns result in “unreliable” conclusions
 - “The City has it backwards.”*

*Garaufis (2009) Memorandum & Order

MQ: College

- Q. Do you know if any steps had been taken at any time to validate that minimum requirement?
- A. [Dr. Schemmer] I do not, sir.

Schemmer, page 325 lines 11-14

MQ

- None of the approaches described in this article by the City’s expert were used:
 - Bobko (2005) Process For Content Validation Of Education And Experienced - Based Minimum Qualifications: An Approach Resulting In Federal Court Approval. *Personnel Psychology*, 58, 771-799.

Bobko on Validity of 7029: P/F

- Q. Is what's in this report, the Bobko, Schemmer report, sufficient to establish that the City's use of Exam 7029 as a pass/fail screening device with a cutoff point of 84.705 is consistent with job relatedness and business necessity?
- A. [Dr. Bobko] No.

(Page 179)

Bobko on Validity of 7029: Ranking

- Q. Is it your opinion that what is in this report is sufficient to establish that the City's rank/order processing and selection of candidates from the Exam 7029 eligibility list is job related and consistent with business necessity?
- A. [Dr. Bobko] No.

(Page 180)

Bobko on Validity of 2043: P/F

- Ditto

IV. Alternative Selection Procedures

- Defense:
 - Oral is impractical
 - Silent on written non-cognitive measures
- Plaintiffs:
 - Other FDs use alternative selection procedures
 - Tenable tests of “softer” areas exist

V. Court Decision

- Background
 - context
 - past litigation
- Standards used by court
- Court decisions in various areas

Context

- NYC population: 52% B/H
- **NYFD: 6% B/H** (as of 1999)
- NYC PD: 48% B/H (in 2010, WSJ 1/7/11)
 - Some exams based on Fleishman areas!
- LAFD: 44% B/H
- Chicago FD: 29% B/H
- Philadelphia FD: 29% B/H

Awareness & Intent

- Plaintiffs arguments cited by court:
 - Long history of adverse impact
 - Continued reliance on similar exams
- Court did not find intent

Similar Past Litigation

- Court described 1973 FD exam case
- City then hired consultants
- City cancelled contracts after 3 years
 - Due to budget crisis

Summary Judgement Standard

- No genuine issue as to any material fact
- Moving party entitled to judgment as a matter of law

Judicial Standard

- Court construes the facts in the light most favorable to the non-moving party
- Court draws all reasonable inferences in favor of non-moving party

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 61

No Triable Issue

- Adverse impact clear
- City's validity evidence is "insufficient as matter of law"

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 62

Assumption that Tests Identify Best Performers

- Cannot rely on this assumption when test is not fair
- "City did not take sufficient measures to ensure that better performers on its examinations would actually be better firefighters."

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 63

Statistical Measures of Adverse Impact

- 80% rule
- Statistical significance

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 64

City's Arguments on Adverse Impact

- Analyses assume perfect parity, which is unrealistic
- Large samples yield erroneous findings
- Rely only on 80% rule

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 65

Court's Response on Sample Size

- "The City has it backwards... large sample sizes make such testing more reliable."

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 66

Court's Findings on Adverse Impact

- Between 10.5 and 33.9 units of standard deviation for P/F
- Between 4.6 and 9.7 units of standard deviation for ranking
- Much greater than 2nd Circuit standard of 2-3 s.d.
- Accuracy of Plaintiffs' statistical calculations undisputed

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 67

Court's Response on "Perfect Parity" Argument

- "...properly assume that racial or ethnic groups will perform equally well...."
- Null hypothesis of no difference is legally appropriate

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 68

Court's Response on "Perfect Parity" Argument (continued)

- Rejected argument of known differences in "capability and preparedness"
- **City needs to show validity to support such an argument**
- Cannot use aptitude disparities in the Adverse Impact stage of case

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 69

Court's Standards for Linkage

- Explain how or why abilities are matched with tasks
- SMES given clear linkage task
- Confirm reliability or agreement of linkage ratings

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 70

Overview of Job Analysis

- About 10 FF/SMEs reviewed existing task list
- 21 Fleishman abilities rated
- Ability to task linkage
 - Linking Panel: 12 FD SMEs

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 71

Court Critique of Test Development

- No review of whether an item measured the intended ability

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 72

Court Relied on Guardians Case

- 1980 case
- 2nd Circuit
- 5 Part Test for Content Validity

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 73

Guardians 5 Part Test for Content Validity

- Suitable job analysis
- Reasonably competent test construction
- Test content must be related to job content
- Test content must be representative of job content
- Scoring system must usefully select those applicants who can better perform job

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 74

Court's Evaluation of 5 Part Test

- For each requirement, the City's arguments were riddled with serious deficits
- "...severe deficiencies at every step..."
- "...impermissibly fail and arbitrarily rank firefighter candidates"

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 75

Job Analysis

- Linkage of abilities to tasks is crucial and flawed
- "...not apparent how they [the tasks] related to the nine specific abilities..."
- SMEs confused about definitions of abilities
- **City ignored "Day One" standard**
 - Tested KSAPs that could be learned on the job

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 76

Questionable Reliance on Landy Study

- Landy study used as a starting point only
- Union resisted Landy's job analysis efforts
- Landy report labeled "Draft"

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 77

Test Questions

- **Developed by FF**, not testing professionals
- No evidence items measured targeted abilities
- Inter-item correlations and factor analysis
 - Will cover this below
- "...some of the cognitive abilities tested are relevant to the job of firefighter"

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 78

Representativeness Requirements

- Test must be a “representative sample of the job”
 - (1) Test content must be representative of job content
 - (2) Test “methodology” must be similar to job procedures
- Reading level should not be pointlessly high

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 79

Test Areas

- Tests ignored most important cognitive abilities (oral comprehension/expression)
- Tests ignored important non-cognitive abilities (e.g., teamwork, dependability)
- “City has no excuse for its failure to test important cognitive and non-cognitive abilities.”

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 80

Reading Level

- Test content should be representative of job content
- More than 50% of questions higher than 12th grade level
- Questioning allowed/encouraged when reading/studying in station and academy
- No questioning allowed during the exam

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 81

Cutoffs

- Guardians warned against setting cut score
 - based on # openings or civil service law (70%)
 - at a point where its unreliability has an extensive impact
- Garaufis: “...no evidence that cutoff bears any relationship to the necessary qualifications...”

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 82

Court on Basis for Cutoffs

- “...using a professional estimate of the requisite abilities...”
- “...or, at the very least...locate a logical break point in the distribution...”
- “...cutoff scores not selected...to measure the minimum...”
- “Nor were they based upon a validity study or job analysis...”

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 83

Plaintiffs’ Critique of Cutoffs

- Bobko/Schemmer report did not even mention cutoffs
- For applicants who took both tests:
 - many passed one and failed the other
 - changed hundreds of ranks

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 84

Court on Rank Ordering

- "...rank ordering...requires strong justification"
- "...Dr. Bobko...stated his report did not establish validity of using rank[s]..."
- Small (chance) changes in scores = large (important) differences in ranks

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 85

City Tried to Put Burden on Plaintiffs

- City: Plaintiffs did not prove test is invalid
- Court: "...City bears burden to show [validity]..."

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 86

Guardians' Cut Score Options

- Three options
- "professional estimate of ... ability levels"
- "a logical 'break point' in the distribution"
- "random ranking of qualified candidates"

- Always need adequate test reliability

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 87

Guardians' Ranking Options

- Adequate job analysis and test construction
- An adequate showing of test reliability
- Random ranking of qualified candidates

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 88

Recap of Some Issues in Case

- Some important abilities not tested:
 - Oral comprehension and expression
 - Non-cognitive abilities
 - Other cognitive abilities
- Test development
 - item content
 - linkage of items to job duties
- Passing points
- Ranking

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 89

Gleanings

- Be sensitive to overall context & EEO posture
- Plan and do work carefully
- Train selection staff
- Document work
- Be intelligent consumers of SME ratings

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 90

Gleanings

- SMEs should not write items
 - Exact role not spelled out in Garaufis Memo
- Need good linkage of abilities to job tasks
- Show items measure intended abilities
- Consider “Day One” needs
 - Avoid testing areas learned on job
- Avoid unnecessarily high reading level

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 91

Gleanings

- Represent important/large parts of job
- Omit major parts of job only with reason
- Defend passing point with job relatedness
 - Especially in a multiple hurdle process
- Reliable test scores

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 92

Gleanings

- Pass point (“Cutpoint”)
 - Basing the passing point for the written exam on the number of openings denied candidates the opportunity to take the second component of the exam
- Perhaps ruling would be different with a one component exam

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 93

Gleanings

- “Big gun” experts should be hired early on
 - Cannot save test after-the-fact

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 94

VI What Happened Next

- Bobko/Cline Exam (“NYC Exam 6019”)
 - Held 2007
- PSI Exam (“NYC Exam 2000”)
 - Held 2012
 - List valid for 4 years

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 95

NYC Exam 6019

- Will not discuss this today
- Many weaknesses
- Not well received by court

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 96

NYC Exam 2000

- 2 1/2 year effort
 - All happened after my involvement ended
- Validation study was filed publicly in 2012
 - I have no confidential information
 - I'll comment on my reading of the report

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 97

NYC Exam 2000

- Teams of experts
 - PSI Team
 - DOJ Project Team
 - Vulcan Society
 - Court-appointed Special Master
- Details of approach

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 98

PSI Team

- John Weiner
- Joseph Abraham
- Sheldon Zedeck
- Donna Denning

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 99

DOJ Project Team

- David Jones
- Leaetta Hough
- Rand Gottschalk

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 100

Vulcan Society Expert

- Harold Goldstein

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 101

Court-Appointed Special Master's Expert

- Shane Pittman

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 102

Overview of Exam 2000

- Computer based test (CBT)
- Cognitive (57 items)
- Non-cognitive (65 items)
- No 80% adverse impact over life of list
 - Reach top 22% of candidates
 - 41,358 candidates (open exam only)

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 103

Cognitive

- Video Lesson
 - Miniature training and evaluation
 - Meta analysis (Robertson & Downs, 1989)
- Operations Manual
 - Work sample

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 104

Cognitive

- Video Lesson (26 items)
 - Lecture on Equipment Operation and Use
 - Narrated slide show on Equipment Safety
 - Student-Instructor Questions & Answers
- Measures
 - Ability to learn and apply information
 - Listening comprehension
 - Reasoning

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 105

Cognitive

- Operations Manual (31 items)
 - Test taker reads excerpt from an operations manual and answers questions
 - Manual page and q on screen simultaneously
- Measures
 - Reading Comprehension
 - Basic Arithmetic
 - Reasoning

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 106

NYC Websites for FF Exam

- Tutorial #1 - Using the CBT System
 - http://www.nyc.gov/html/dcas/html/work/firefighter_cbt-tutorials_1.shtml
- Sample questions (21 minutes)
 - http://www.nyc.gov/html/dcas/html/work/firefighter_cbt-tutorials_2.shtml

(URLs last accessed on 5/18/2014)

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 107

Non-Cognitive

- Background questions
 - Dependability
 - Activity
 - Agreeableness
 - Even Tempered
 - Low Anxiety
 - Self-Esteem

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 108

Non-Cognitive

- Assessment dimensions
 - Conscientiousness
 - Agreeableness
 - Emotional Stability
 - Interpersonal Competence
 - Honesty/Socialization

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 109

Weights: Exam 2000

- 50%-50% Cognitive and non-cognitive
- Cognitive component weights
 - 70% video lesson
 - 30% operations manual
- Non-Cognitive
 - Equally weighted areas

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 110

Validation Efforts

- Content
- Criterion-related
 - Academy (N=446)
 - Job Performance (N=566)
- Construct

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 111

Validity Report Strengths

- Content validation study
 - Readability analyses of job materials
 - Linked abilities to task categories
- Criterion-related validation studies
 - Academy
 - Job performance
 - Both Ns about 500

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 112

Validity Report Strengths

- Construct validation study
- Fairness study
 - No under-prediction for minority firefighters
- Scoring process determined before the exam

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 113

Validity Report Strengths

- No adverse impact using 80% rule for any of the 4 years the test is used
- AI ratios reported as: (**note signs of sd's!**)
 - .983 for year 1, -2.26 S.D.
 - .912 for year 2, -1.64 S.D. (not significant)
 - 1.061 for year 3, 1.67 S.D. (not significant)
 - 1.037 for year 4, 1.44 S.D. (not significant)

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 114

Other Features

- Mean difficulty of cognitive items = .7
 - S.D. = .18
- Mean point-biserial for cog. items = .4
 - S.D. = .09
- Transformed CBT scores to 0 to 100 scale
 - mean = 88
 - median = 92, after special (e.g. resident) points

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 115

Other Features

- Computer administered
 - items not in same order for all booklets
- Multiple forms of cognitive test
- Multimedia
 - video
 - audio
 - graphic images

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 116

Other Features

- Several types of response format
 - single answer
 - multiple answers
 - drag and drop (into correct order)
 - graphic stems and response options
- Promotional exam for 873 EMTs

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 117

Validity Report Wrinkles

- Some surprising lack of content
 - Not there for a reason?
 - Perhaps audience is court not I/O psychologists
- Some surprising content
- Strong caveat: I may have missed or misinterpreted some points in the PSI test development and validation report.

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 118

Validity Report Wrinkles

- No discussion of why AI is low
- No report of d for non-minority vs. minority
- No report of mean and S.D. by racial/ethnic group

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 119

Why Low Adverse Impact?

- No comparison of d of 2 cognitive tests
 - Video Lesson
 - Operations Manual

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 120

Validity Report Wrinkles

- No comparison of *d* of various types of cognitive items
 - Single correct choice
 - Multiple correct choice
 - Orally presented
 - Graphic content

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 121

Validity Report Wrinkles

- Reliability of job analysis means, $r = .99+$
 - For each rating scale
- Reliability of job performance, $.21 < r < .37$
 - Reports corrected validity correlations only based on these low criterion reliabilities
 - Reliability of communication criterion measure called “very low”
 - Should this have triggered more research?

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 122

Validity Report Wrinkles

- Level of criterion-related validation
 - $r = .25^*$ with **academy** performance (N=446)
 - $r = .12^{**}$ with **job** performance (N=566)
 - *estimated from corrected r of .30
 - **estimated from corrected r of .25
- How impressive is predicting academy m/c test scores?

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 123

Validity Report Wrinkles

- Correlation: Academy & Job Performance
 - 48 correlations (quizzes, midterm, final, exercises)
 - $.06 < r < .21$ (uncorrected)
 - Fisher Z average of all 28 correlations is **.11**
- Academy weakly predicts job performance
- Product of correlations: $.25 * .11 = .03$

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 124

Validity Report Wrinkles

- Total score precision of measurement, $\alpha = .88$
- Alternate CBT Forms correlate with validated Form A, $r \geq .87$, uncorrected*
- But $.88$ squared is $.77$

*Did not say “corrected”

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 125

Validity Report Wrinkles

- Task criterion: 2/3 rated task as important or very important
- But two types of fire companies:
 - Engine, with 62% of job analysis sample
 - Ladder, 38% of job analysis sample
- Highest S.D. for engine and ladder tasks
- Did 2/3 criterion eliminate much of FF job?

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 126

Validity Report Wrinkles

- Reading level of exam reflects range of job
 - Rather than minimum level for non-reading items
- Construct validity of cognitive areas average of .19 (uncorrected)
 - 10 cognitive components with 4 EAS tests
 - 40 correlations

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 127

Validity Report Wrinkles

- No formula given for overall score
 - Weighted 50-50, but did they standardize?
- No report of number certified each year, or overall, by racial/ethnic group

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 128

Validity Report Wrinkles

- Job analysis “Quality Check Items”
 - 5 tasks and 5 abilities clearly not job related
- Surveys excluded if at least **3** tasks rated:
 - Performed: Yes and Importance > 2
 - Frequency > 2
- Surveys excluded if at least **3** abilities rated:
 - Important >2 and Day 1 Yes
- Too lenient!

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 129

Why Low Adverse Impact?

- Minimum qualifications
- HS diploma or GED
- Plus **one** of the following
 - 15 college semester credits
 - honorable discharge from military
 - 6 months paid work experience
- Applicants above average, on average

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 130

Why Low Adverse Impact?

- Practice test video
 - better prepared for exam, less test anxiety
 - less stereotype threat (minority candidates in video)
 - may have caused self-selection
- Question types
 - Select several correct answers (M/C)

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 131

Why Low Adverse Impact?

- Low passing point
 - 97.7% pass open exam
 - 97% of minorities pass
 - P/F AI at passing point (means not reported)
 - “Low” AI at effective passing point
- Passing point set 2 SE below the point that predicts “just adequate” performance

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 132

Why Low Adverse Impact?

- Redefined adverse impact of exam
 - Each year's adverse impact ratio ignores applicants considered in previous years.
 - No report of the number of minority candidates
 - at each score point
 - at each year's effective pass point
- Adverse impact at pass point highly signif.

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 133

Why Low Adverse Impact?

- Cognitive test based mainly on 26 items
 - But 26 items may have reliability of .7
- A few items were deleted
 - But not from all forms of the test

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 134

Why Low Adverse Impact?

- Promotional list exhausted before open used
 - 873 EMTs and Paramedics
 - Perhaps minorities are over-represented
- Perhaps residency and veterans points
 - Special points added 3.2 points per passer
 - 94% of special points were for NYC residency

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 135

Why Low Adverse Impact?

- Decompose mean CBT score
 - 100 is mean raw score

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 136

Why Low Adverse Impact?

- Decompose mean CBT score of 88
 - mean score on cognitive = $.69 \times 57 = 38.4$
(mean item difficulty * number of items)
 - $88 = 38.4 + \text{mean for non-cognitive}$
 - mean for non-cognitive = 49.6
 - Perhaps weighted without standardization
 - Perhaps non-cognitive drove rankings more than 50% weight

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 137

Why Low Adverse Impact?

- Cognitive test measures 13 abilities
 - Job analysis questionnaire had 89 abilities
 - I did not look at the mean ability ratings
- Perhaps tested more memory, less other cognitive areas
 - Memory has lower AI

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 138

Why Low Adverse Impact?

- If NYC is basically ranking on personality, greater validity might be available by using the Greatest Strength Method (Wiesen & Aguinis, 2010)

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 139

Questions on PSI Test

- Does the PSI test push the field forward?
- If so, how?
- If not, why the low adverse impact with certification of the top 20% of test takers?

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 140

VII Gleanings

- Context important
 - NYPD vs. NYFD EEO posture
 - History of NYFD's legal involvement
- Adverse impact is huge
 - Hard to overcome a shockingly small AI ratio

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 141

VII Gleanings

- Focus on non-cognitive to reduce AI
- Collaborative relationships w plaintiffs
 - Agreements rather than confrontations

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 142

References

- Garaufis, N. G. (2009). *Memorandum & Order in USA and Vulcan Society, et al. vs. City of New York, et al., July 22, 2009*. United States District Court; Eastern District of New York. Retrieved May 20, 2014 from <http://www.clearinghouse.net/chDocs/public/EE-NY-0192-0008.pdf>.
- El-Ghobashy, T. (2011). Minorities Gain in NYPD Ranks. *Wall Street Journal*, 1/7/2011.

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 143

References

- PSI Services LLC (2012). *FDNY Firefighter Test Development and Validation Report*. Author: Burbank, CA. (Case 1:07-cv-02067-NGG-RLM Document 976-1 Filed 09/24/12.)
- Robertson, I. S. & Downs, S. (1989) Work sample tests of trainability: a meta-analysis. *Journal of Applied Psychology*, 74, 402-410.

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 144

References

- *Sample Questions*. Retrieved May 18, 2014 from http://www.nyc.gov/html/dcas/html/work/firefighter_cbt-tutorials_2.shtml
- *Tutorial #1 - Using the CBT System*. Retrieved May 18, 2014 from http://www.nyc.gov/html/dcas/html/work/firefighter_cbt-tutorials_1.shtml

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 145

References

- Wiesen, J. P. & Aguinis, H. (2010). *New Methods for Reducing Adverse Impact and Preserving Validity*. Presentation at the annual conference of the society for Industrial and Organizational Psychology, Atlanta, GA.

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 146

Your Questions/Observations

Copies of this presentation are available at:
<http://appliedpersonnelresearch.com/pubs.html>

<http://appliedpersonnelresearch.com>

Wiesen (2014), PTC/MW May Luncheon 147