

# **The Recent Judicial Decision Concerning Two NYC Firefighter Exams**

Joel P. Wiesen, Ph.D.

wiesen@appliedpersonnelresearch.com

MAPAC Fall Conference, 11/5/09

Albany, NY

# Today's Presentation

- I Overview of Case
- II Adverse Impact
- III Validity
- IV Alternative Selection Procedures
- V Court Decision

# Today's Goals

- Learn from Court's written decision
- Gain insights from case facts that are not in the decision

# I. Overview of Case

- Snapshot of case
- “Typical” Title VII Impact Case
  - But two Plaintiffs
- Job of Firefighter
- Decided on Summary Judgement

# Snapshot of Case

- Challenge to NYC Firefighter entry exam
- Minimum qualifications, written test, PPT
- Case brought by DOJ
- Vulcans Society intervened

# “Typical” Title VII Impact Case

- Adverse Impact
- Validity
- Alternative Selection Procedures

# Two Plaintiffs

- US DOJ
- Vulcans (intervenors)

# Job of Firefighter

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



# Decided on Summary Judgement

- Unusual
- We will review court's reasoning later

## II. Adverse Impact

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

# Minimum Qualifications (MQs)

- 30 college credits (any courses)
- Driver's License
- Certified First Responder Certification with Defibrillation
- No validation of these MQs

# 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.”

# Driver's License

- Adverse Impact
- Validity
  - Driving task dropped out in job analysis.
  - Most Firefighters do not drive on the job
  - Chauffeur is a promotional title with extra duties and pay.

# Certified First Responder Certification with Defibrillation

- 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



# City's Validation Reports

- One test development report
  - Mainly job analysis and test plan
- First exam: 9 page report with 9 appendices
- Second exam: “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)

# Major Testing Flaws

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

# Test Design Flaws

- Aspects of the Physical Performance Test (PPT)
  - Will not discuss today
- Use of Fleishman abilities
- Limiting KSAPs to Fleishman abilities
  - SMEs could not add to ability list

## Fleishman Abilities Not Understood

- “Linking Panel” SMEs did linkages and rated abilities
  - 3/4 of these SMEs did not understand all the Fleishman abilities
- Examiners who wrote questions also did not fully understand the Fleishman abilities

# 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 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.



# 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."

# 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

# 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)

# Test Development Flaws

- Job analysis
- Test questions

# Job Analysis

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

# 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

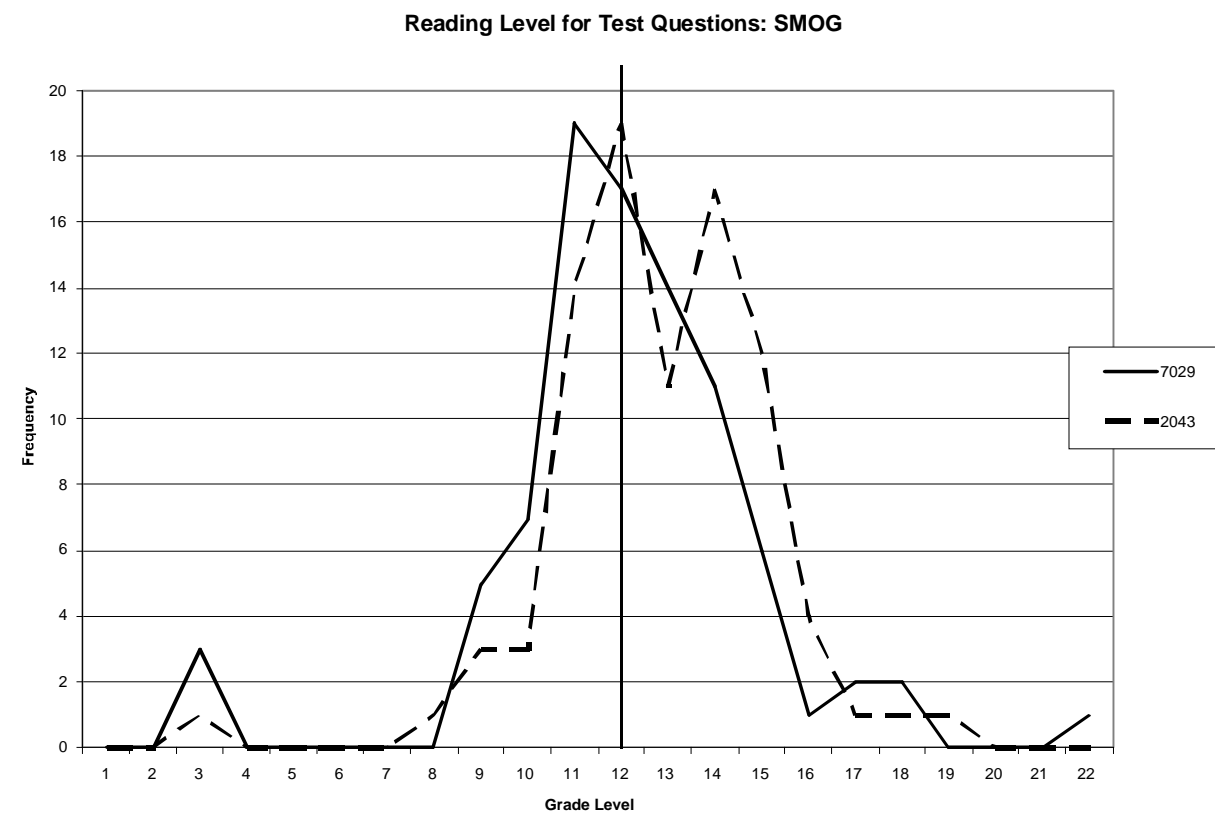
- Link to test areas
- Reading level
- Factor structure

# 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

# Wording and Adverse Impact

- Longer questions harder
  - Greater effect for blacks
- Questions with jargon harder
  - Greater effect for blacks

# Factor Structure

- Factor analysis yielded 1 main factor and a possible second that did not line up with any intended ability.
- Inter-item correlations between abilities higher than within.

# Measuring Unintended Abilities

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

# Some Topics Not Covered Today

- PPT
  - Design
  - Implementation

# Review 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, cut point or 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

# 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

## 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 impractical
  - Silent on written non-cognitive measures
- Plaintiffs:
  - Other FDs use alternative selection procedures
  - 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
- LAFD: 44% B/H
- Chicago FD: 29% B/H
- Philadelphia FD: 29% B/H

# Awareness = Intent

- Argument cited by court:
  - Long history of Adverse Impact
  - Continued reliance on similar exams

## 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

# No Triable Issue

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

# 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.”

# Statistical Measures of Adverse Impact

- 80% rule
- Statistical significance
- Both accepted by court

# Four Challenged Practices

- P/F of two exams
- Ranking of two exams

# City's Arguments on Adverse Impact

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

# Court's Response on Sample Size

- “The City has it backwards... large sample sizes make such [statistical] testing more reliable.”

## 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



## 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

## 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

# Overview of Job Analysis

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

# Court Critique of Job Analysis

- No explanation in report of reason for links between abilities and tasks
- No study of reliability of job analysis ratings

# Court Critique of Test Development

- No review of whether an item measured the intended ability

# Court Relied on Guardians Case

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

# 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 from among applicants who can better perform job

## 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."



# Job Analysis

- Linkage of abilities to tasks is crucial and flawed
- “...not apparent how they [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

## Unclear Reliance on Landy Study

- Landy study used as a starting point only
- Did not have a test development report describing role of Landy's past work.
- Landy's previous work not submitted to court.
- Union resisted Landy's job analysis efforts

# Court Critique of Test Questions

- Developed by FF, not testing professionals
- No evidence items measured intended abilities
- Inter-item correlations
- Factor analysis
- :...some of the cognitive abilities tested are relevant to the job of firefighter.”

# Representativeness Requirement

- 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

# Court Critique of Test Areas

- Tests ignored most important cognitive abilities
- Tests ignored important non-cognitive abilities
- “City has no excuse for its failure to test important cognitive and non-cognitive abilities.”

# 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

# 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
- Here “...no evidence that cutoff bears any relationship to the necessary qualifications...”

## 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....”



# 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
  - dual passers changed hundreds of ranks

# 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.

# City Tried to Put Burden on Plaintiffs

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

# Guardians' Cut Score Options

- adequate job analysis and test construction
- adequate reliability
- random ranking of qualified candidates

# 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

# Summary

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

# Your Questions/Observations

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