

Disparate Impact and Employment Testing: A Legal Update

Michael M. Harris, Ph.D.

University of Missouri-St. Louis

(mharris@umsl.edu)

EASI Consult LLC

(mharris@easiconsult.com)

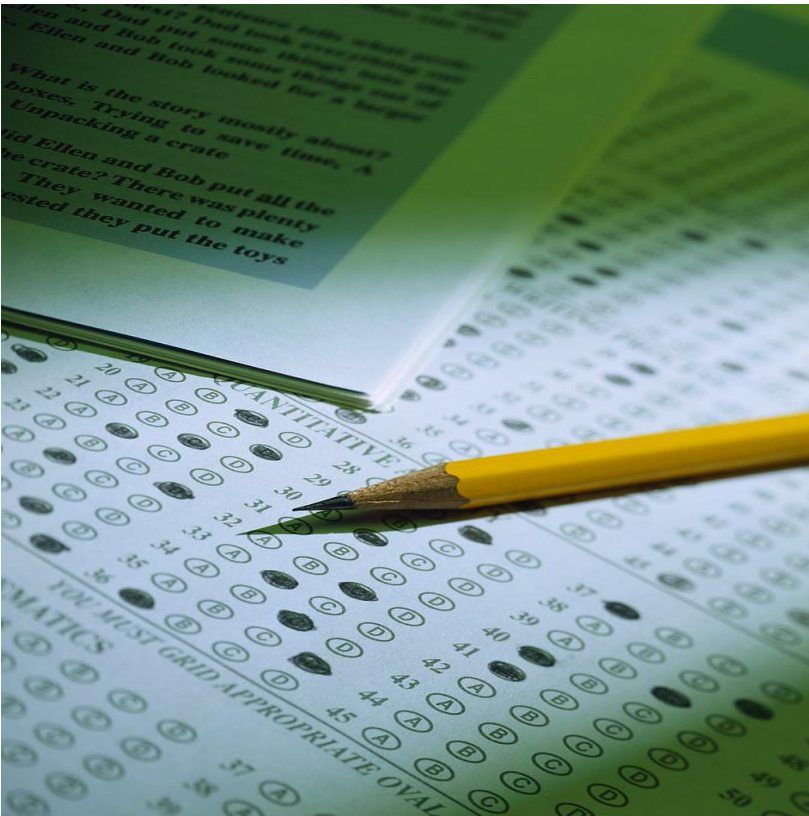
Proof in Adverse (Disparate) Impact Cases



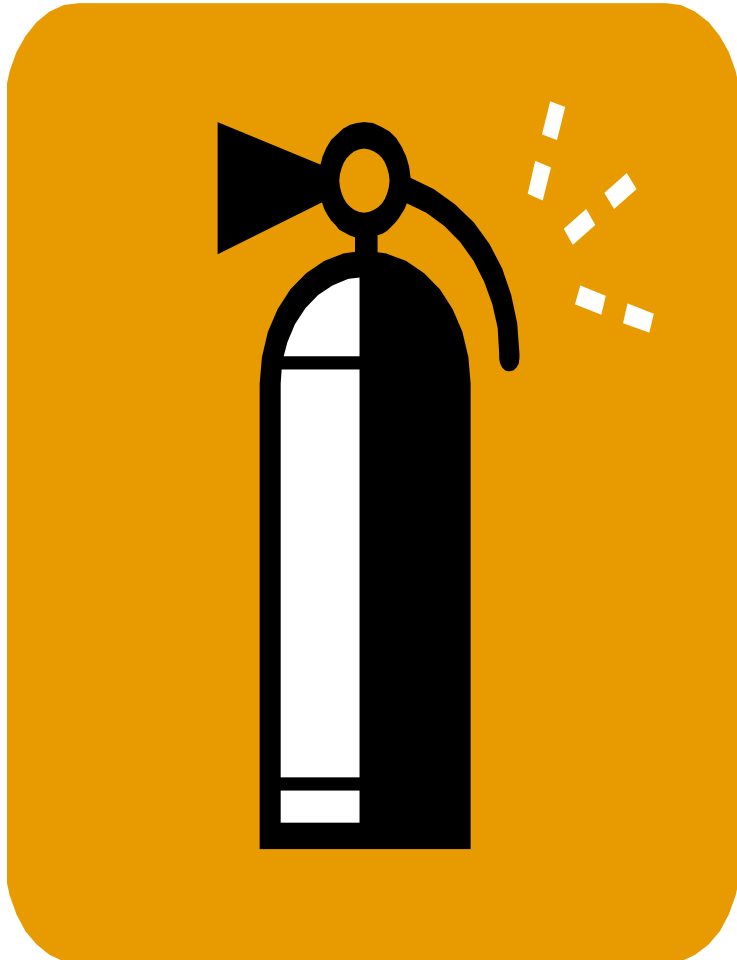
- Prima facie case
- Job-related and consistent with business necessity
- Alternative practice with lesser impact

Prima facie case

- Typically, the 4/5's rule is used
 - But this rule is not always accepted by the courts
- Recent cases
 - Isabel v. City of Memphis
 - Based on 4/5's rule, no adverse impact (83%)
 - Based on cutoff score on written test, substantial disparity in test scores for Blacks and Whites



Prima facie case (cont.)



- Stewart v. City of St. Louis
- Firefighter promotion exams across several years
- On 2000 exam for Battalion chiefs, 16 Whites and 10 Black applicants passed, but only 2 were promoted (both Whites)
- Judge ruled numbers were too small to use 4/5's rule; also, statistical analyses showed no significant differences
- For 2000 Fire Captain exam, defendant's expert calculated 4/5's rule based on number of promotions compared to numbers who passed the exam
- Judge ruled that 8th circuit had previously used the total applicant pool and based on this, there was adverse impact here

Prima facie case (cont.)



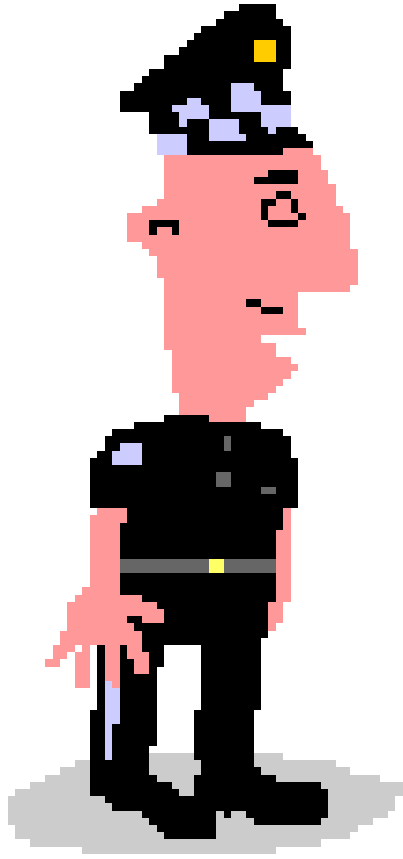
- Reynolds v. Alabama Dept. of Transportation (2003)
 - Plaintiffs argued that use of unit weighted tests created adverse impact, based on racial differences on exams
 - Defendant argued that this is not evidence it has led to actual disparities in hiring
 - Court favored defendant's argument, despite the argument by plaintiffs that their position followed the Teal case
 - Because the tests here were used to rank order candidates, not a knock-out factor as in Teal

Job-related and consistent with business necessity



- **Validity Evidence**
- Many of the cases involve content validity
- But there were a few cases involving criterion-related validity

Criterion-related validity evidence



- U.S. v. Delaware (2004) involved careful examination of criterion-related validity results
- Concurrent design
- Alert test
 - 160 item test measuring reading and writing
- 62 supervisors rated their employees on 13 dimensions
 - PDRF composite of 4 dimensions

U.S. v. Delaware (cont.)



- Uncorrected correlations ranged between .15 - .24; all statistically significant
 - When corrected for range restriction and unreliability, correlations ranged between .24 - .39
 - But no statistical significance tests provided

U.S. v. Delaware (cont.)



- Judge concluded that the “Alert has generally low criterion validity...” because none of them is .3 or higher
- Therefore, he argued that such tests must be used with “particular care”

Criterion-related validity evidence



- **Bradley v. City of Lynn**
- Entry-level firefighters
- Used written cognitive ability test
- Citing expert witness testimony that “Correlation coefficients around .30 -- .40 are considered acceptable” and that a rule of thumb of .3 should be the minimum correlation
- In addition to other problems, the C-R validity for the test seemed to be between .20 and .30

Content Validity

- Given that many of the cases involve safety (i.e., police, fire) positions, and especially promotions, it is not surprising that content validity is a popular validation strategy

Stewart v. City of St. Louis



- Plaintiff's expert argued that consultants did an inadequate job analysis, because
 - 1. They used a “semi-structured interview,” not a questionnaire

Stewart v. City of St. Louis

- Without questionnaire, expert asserted there was no “paper trail”
- Defendant’s experts argued that this approach did not lessen its validity
- 2. Vast majority of questions on the job knowledge test focused on 2 major work behaviors

Stewart v. City of St. Louis



- Defendant's experts countered that focus on questions related to 2 major behaviors will increase prediction of job success more than other focusing on lower-ranked behavior
- 3. Insufficient documentation in validity reports to make determinations and therefore tests invalid



- Judge noted that Guidelines don't specify that everything has to be included in a report
- Also, that plaintiffs' expert never indicates how much would be enough – he fails to set forth his own methodology or standards for his opinion

- 4. Tests should have been reviewed by Fire Department staff prior to use
- Defendant's expert responds that this doesn't render tests invalid

Hearn v. City of Jackson (MS)



- Police promotions
- In 1998, consulting firm developed a 3 stage test (written, AC, panel interview)
- Plaintiffs claimed that C-R validity should have been used, not content validity

Hearn v. City of Jackson (MS)



- Defendant's expert said if possible Content validity would have been done, but it was not feasible
- Plaintiffs argued that written test makes no sense as information could be looked up on the job
- Judge said that test measured what officers should readily know
- Plaintiffs argued that job analysis was done several years earlier (1993)
 - Judge says no requirement to have new job analysis assuming it stays relevant and accurate
 - Consultant confirmed with SMEs that it was still accurate
 - Another expert said that a new job analysis is needed only if department has changed structurally or if positions have been added or eliminated, thus changing the job

Hearn v. City of Jackson (MS)

- Insufficient number of SMEs
 - Judge says all sergeants were used and even some higher positions (SMEs=7)
- SME rating system not sufficiently defined
- Linkage of tasks and KSAs not done on timely basis (i.e., it was done before test was prepared)
 - Judge says this was done at or before test construction, is acceptable
- Internal reliability too low (.79)
 - Defense says .79 is not too low, esp. for homogenous test
- Test had poor readability and expert felt higher reading level than required for the job
 - Defense expert argued that reading level required by job was much higher than plaintiffs expert says

Focusing Too Highly On A Few KSAs May Not Be Good, Either

- Bradley v. City of Lynn
 - Judge concludes that there is no evidence that cognitive ability “as the sole basis” for rank ordering fire fighters is valid.
 - She argues that Teamwork and physical prowess are even more correlated with job performance

Content vs. Criterion-Related Validity



- Stewart v. City of St. Louis
- Plaintiffs' expert says C-R would have been more appropriate
- Defense expert
 - has never seen C-R obtained for promotional testing
 - Test security concerns
 - Sample size problems
 - Test-taker motivation

Alternative practice with lesser impact



- Bradley v. City of Lynn
- Banding
- Other tests (e.g., personality, biodata)
- Judge says that the plaintiffs need not “provide the exact floor plan”
 - Don’t have to have the exact items or tests

Not Just the Public Sector (But, Mostly)



- Perhaps because of the need for custom-tailoring
- Some large class certification cases involve tests (e.g., FedEx)
- Apprenticeship programs

Conclusions



- Judges are increasingly sophisticated about disparate impact, validity, and tests in general
- Don't wait until litigation to do validation studies (“studies done in anticipation of litigation to validate discriminatory employment tests...must be examined with great care due to the danger of lack of objectivity”)

Conclusions (cont.)



- Don't take the search for alternatives lightly
- Test construction documentation is critical to a case, including dates
- Stay current with law cases
 - EEOC's Employment Testing and Screening Meeting (esp. L. Ashe's summary of legal cases)

Cautions



- Different judges
- Different circuits
- Why litigation?

Questions??