

Statistical Support of Test Fairness Reconsidered

29th Annual IPMAAC Conference
Orlando, Florida; June, 2005

Joel P. Wiesen, Ph.D.

Applied Personnel Research

Wiesen@PersonnelSelection.com

Why Ask About Test Fairness?

- We know there are many reasons for the adverse impact seen with our tests.
- Are our tests one of the reasons or one of the symptoms?

Accepted Wisdom: Tests Are Fair

- Tests of cognitive ability are valid for virtually all jobs (Schmidt, 1988)
- Job performance measures are unbiased (Schmidt, 1988)
- The issue of test bias is scientifically dead (Hunter & Schmidt, 2000)

Some Indications Tests May Not Be Completely Fair

- Test vs Job Performance Differences
- Sternberg's multicultural work
- Critiques of educational tests
- Psychometrics and restrictive immigration law

Test vs Job Performance Differences

- Big, 1 s.d., difference in B-W mean test scores
- Smaller, .5 s.d., difference in B-W mean job performance
- Is this fair?

Prediction Equation

- $y = r * x$
- y is job performance
- x is test performance
- r is validity coefficient

Sternberg's Multicultural Work

- In some cultures, what you learn outside the classroom determines success in life
- Culture free tests are not (Sternberg, 2004)

Critiques of Educational Tests

- ...test publishers...assess the...validity...by correlating...with other such tests or with external criteria, most typically school grades
- ...and that can make tests look good, regardless of whether they are” (Sternberg, 2000, pg 159 and 160).

Psychometrics and US Restrictive Immigration Law

- In early 1900's Goddard said 79% of immigrant Italians were “feeble-minded”
- Today Italian American students show slightly above average IQ
- What was taken as hereditary was just lack of familiarity with the English language and our American culture (Sternberg, 2004)

Reasons to Question Our Validity Studies?

- Limited types of criterion measures
- Indications of race-related bias in criteria

Ideal Criterion Measure

- “In criterion-related validation, the criterion should be reliable, valid and free from third-variable biases. It is amazing how easy that sentence is to write and how difficult it is to accomplish.” (Guion, 1998, page 446)

Limited Criterion Measures

- Training scores, e.g., Police Academy
- Supervisor evaluations

Validity of Training Scores

- Method bias if based largely on M/C tests
- Higher validity than job performance:
 .76 vs .38 for law enforcement (Hirsh,
 Northrup & Schmidt, 1986, as cited in Schmidt and
 Hunter, 2004)
- Learning faster versus doing the job

Research Related to Supervisor Evaluations

- Height bias
- Attractiveness bias
- Unconscious race bias
- Low correlation with other measures of
job performance

Height Bias

- Each inch corresponds to \$789 in additional annual income, on average (Judge & Cable, 2004)

Attractiveness Bias

- Average effect size of .37
- Equally important for men and women
- Professionals as susceptible to this bias as college students (Hosoda et al., 2003).

Unconscious Race Bias

- People with black-appearing facial features are judged in stereotypical (negative) fashion
- People judging are not aware of this. They think they are relying on facial expressions and the like (Blair, et al., 2002 and Wittenbrink, et al., 1997).

Studies of Race of Supervisor and Employee

- Black and white supervisors rated white employees about the same
- White supervisors rated black employees much lower than black supervisors did.
- Mistaken belief: Supervisory ratings are not biased on the basis of race (Stauffer & Buckley, 2005)

Supervisor Evaluations and Other Measures of Job Performance

- Mean correlations of .25 between hands-on tests and supervisor ratings
- Mean correlations of .5 between paper/pencil job knowledge tests and supervisor ratings (Wigor & Green, 1991, as cited by Wagner, 1997).
- This raises, again, the criterion question

In Conclusion

- The evidence for test fairness may not be as strong as some think
- Tests are not as effective as we would like
- Yes, there are B-W differences in schooling
- We should continue to strive to understand and reduce adverse impact on our tests
- Innovative R&D needed

R&D Thoughts

- Job analysis: include wider range of KSAPs (e.g., interpersonal), look more closely at tasks
- Job analysis: include HR and I/O w/ SMEs
- Job criteria: more multidimensional
- Tests: new test modes/question types
- More research cooperation with colleges

References

- Blair, I.V., Judd, C.M., Sadler, M.S., and Jenkins, C. (2002) Role of afrocentric features in person perception: judging by features and categories, *Journal of Personality and Social Psychology*, 83, 5-25.
- Guion, R.M. (1998) *Assessment, Measurement, and Prediction for Personnel Decisions*. Mahway, New Jersey: Lawrence Erlbaum Publishers.
- Hunter, J.E. & Schmidt, F.L. (2000) Racial and Gender Bias in Ability and Achievement Tests. *Psychology, Public Policy, and Law*, 6, 151-158.
- Hosoda, M., Stone-Romero, E.F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: A meta-analysis of experimental studies. *Personnel Psychology*, 56, 431-462.
- Judge, T.A. & Cable, D.M. (2004) The Effect of physical height on workplace success and income. *Journal of Applied Psychology*, 89, 428-441.

References (continued)

- Schmidt, F.L. (1988) The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior*, 33, 272-292.
- Schmidt, F.L. & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, 86, 162-173.
- Stauffer, J.M. & Buckley, M. R. (2005) The Existence and Nature of Racial Bias in Supervisory Ratings. *Journal of Applied Psychology*, 90, 586-591.
- Sternberg, R.L. (2000) Implicit theories of intelligence as exemplar stories of success: Why intelligence test validity is in the eye of the beholder. *Psychology, Public Policy, and Law*, 6, 159-167.
- Sternberg, R.L. (2004) Culture and Intelligence. *American Psychologist*, 59, 325-338.

References (continued)

- Wagner, R.K. (1997) Intelligence, Training, and Employment. *American Psychologist*, 52, 1059-1069.
- Wittenbrink, B, Judd, C.M., and Park, B. (1997) Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures, *Journal of Personality and Social Psychology*, 72, 262-274.