Improving the Validity & Fairness of a Firefighter Selection Process with the addition of Integrity Testing

Industrial/Organizational Solutions, Inc.

July 23, 2012

Presented at the 2012 conference of the International Personnel Assessment Council

David L. Blitz, Ph.D. & **John V. Harnisher, Ph.D**.

Introduction

✦ Cognitive ability is one of the best predictors of job performance (e.g., Hunter & Hunter, 1984; Jensen, 1980; Landy, Shankster & Kohler, 1994; Schmidt, Ones & Hunter, 1992)

✦ Cognitive-based selection instruments have tended to adversely impact protected groups (e.g., Cleary, Humphreys, Kendrick & Wesman, 1975; Hartigan & Wigdor, 1989)

✦ Selection researchers have turned to non-cognitive measures as supplements or alternatives (e.g., Barrick & Mount, 1991; McHenry, Hough, Toquam, Hanson & Ashworth, 1990)

✦ Personality-based measures (e.g., "Big 5", honesty/integrity) have been shown to be predictive of job performance (e.g., Barrick & Mount, 1991; Ones, Viswesvaran & Schmidt, 1993) and can add incremental validity (McHenry et al., 1990; Schmidt & Hunter, 1998)

✦ Structured interviews have also been shown to add incremental validity (e.g., Schmidt & Hunter, 1998)

Background

- Fire department for a mid-sized city in the Eastern U.S.

- Concerned about adverse impact with strict cognitive-based testing

- Concerned about finding and selecting candidates who will be able to perform the job well

- Dedicated recruiting efforts

- Civil service regulations:

  - Codified passing point (75)

  - Top-down selection only (banding NOT an option)

  - 2-year eligibility list

  - Stated preference for residents and military veterans

Desired Goals

- Reduced adverse impact across protected groups

- Increased selection of residents

- Increased selection of military veterans

- Sufficient candidate pool to be able to fill 2-3 cadet classes (i.e., approximately top 140-150 from eligibility list)

Methodology

- Six-month advance public notice
  - Exam components
  - Hurdles
  - Passing standard
  - Residency/military service preferences
  - Candidate review process
- Written exam administered first
  - Instrument 1: cognitive and general personality measures
  - Instrument 2: integrity measure
  - Written exam composite score: 20% cognitive, 20% personality, 60% integrity (weights set *a priori*)

Methodology (cont'd)

- Structured Oral Interview (SOI) administered next
    - Not possible to administer SOI to all candidates for practical/logistic reasons
    - Candidates with a written exam composite score greater than or equal to 75 were invited to be interviewed
    - Combined composite score calculated (40% written exam, 60% SOI; weights set *a priori*)
- Preference points added for candidates with combined composite score greater than or equal to 75
- Final scores calculated and initial eligibility list created
- Response/scoring reviews conducted for those candidates requesting a review
- Eligibility list finalized

Evidence of Validity

- Three-pronged approach:
    - Concurrent validation study
    - Predictive validation study
    - Validity transportability study

- Concurrent validity study
    - Insufficient data collected at time of administration to incumbents
    - Will assess if further data collection is feasible

- Predictive validity study
    - Criterion data will be collected when new hires graduate from academy

- Transportability study
    - 100% task overlap between source and target agencies; consistent with Gibson & Caplinger (2007), permissible to transport validity
    - Correlations (uncorrected) with job performance:

    Cognitive Ability   .21
    Personality   .47
    Integrity   .28

Adverse Impact (AI) Ratios

- The following tables present Adverse Impact Ratios (AIRs) by Ethnicity based on the (approximately) top 140-150 candidates

- Attention is focused on this group because, for all practical purposes, selection will not likely be made down to the passing point

- Baseline groups for AIR calculations are Caucasians and Males for Ethnicity and Gender, respectively; while this deviates from the *Uniform Guidelines*, it is done for illustrative purposes to highlight the comparisons with the two groups most commonly of interest

AI Ratios (cognitive and personality)

Ethnicity	N (fail)	N (pass)	SR	AIR
Not Reported	14	2	12.5	0.48
African American	118	6	4.8	0.18
Asian/Pacific Islander	3	0	0.0	0.00
Hispanic	119	14	10.5	.40
Native American	3	0	0.0	0.00
Caucasian	340	121	26.2	Baseline
Other	26	1	3.7	.14
Gender	N (fail)	N (pass)	SR	AIR
Not Reported	6	3	33.3	1.79
Male	612	140	18.6	Baseline
Female	5	1	16.7	.90

AI Ratios (integrity added)

Ethnicity	N (fail)	N (pass)	SR	AIR
Not Reported	12	4	25.0	1.19
African American	108	16	12.9	0.61
Asian/Pacific Islander	3	0	0.0	0.00
Hispanic	112	21	15.8	.75
Native American	2	1	33.3	1.59
Caucasian	364	97	21.0	Baseline
Other	24	3	11.1	.53
Gender	N (fail)	N (pass)	SR	AIR
Not Reported	7	2	22.2	1.22
Male	615	137	18.2	Baseline
Female	3	3	50.0	2.75

AI Ratios (SOI added)

Ethnicity	N (fail)	N (pass)	SR	AIR
Not Reported	0	0	N/A	N/A
African American	113	19	14.4	0.66
Asian/Pacific Islander	4	2	33.3	1.53
Hispanic	125	15	10.7	.49
Native American	2	0	0.0	0.00
Caucasian	377	105	21.8	Baseline
Other	4	1	20.0	.92
Gender	N (fail)	N (pass)	SR	AIR
Not Reported	0	0	N/A	N/A
Male	621	140	18.4	Baseline
Female	4	2	33.3	1.81

AI Ratios (preference points added)

Ethnicity	N (fail)	N (pass)	SR	AIR
Not Reported	0	0	N/A	N/A
African American	103	29	22.0	1.19
Asian/Pacific Islander	4	2	33.3	1.80
Hispanic	119	21	15.0	.81
Native American	2	0	0.0	0.00
Caucasian	393	89	18.5	Baseline
Other	4	1	20.0	1.08
Gender	N (fail)	N (pass)	SR	AIR
Not Reported	0	0	N/A	N/A
Male	620	141	18.5	Baseline
Female	5	1	16.7	0.90

Summary: Validity

- Initial efforts to collect concurrent sample data insufficient

  - Not sure if this was due to mistrust, workload, or both
  - Efforts to collect additional data may continue

- Collection of predictive sample data yet to begin

  - Next Academy class yet to be selected

- Validity transport successful

  - Serves as baseline only

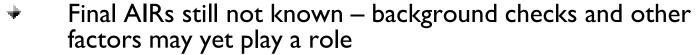- Not yet able to determine if approach led to *improved* validity

Summary: Fairness

- Cognitive and Personality measure
  - Provided baseline – room for improvement

- Adding the integrity measure
  - Definite improvement in terms of fairness

- Adding the structured oral interview
  - May not have added much in terms of improved fairness

- Adding preference points
  - Definite improvement in terms of fairness
  - Meets desired goal of increased resident/military representation

Conclusions

- Final AIRs still not known – background checks and other factors may yet play a role
- Nothing new in terms of science or practice – this was more a case study of applying what we know about selection testing
- Adding multiple non-cognitive measures to the cognitive measure was beneficial, as expected – *a priori* weights did not completely "work" as expected but can be refined in future
- Small changes in candidate pass/fail resulted in sometimes larger shifts in AIRs, underscoring much of what's been voiced regarding dissatisfaction with the 4/5ths Rule
- Statistics presented are "simplistic" – effect size measures and significance tests will add more to the picture
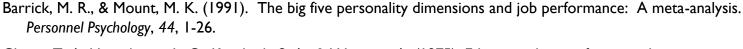- Very few candidate reviews requested – procedural justice at work?

For Further Information

To obtain further information please contact us at:

Industrial/Organizational Solutions, Inc.
Attn: Research & Development Department
1127 S. Mannheim Rd., Ste. 203
Westchester, IL 60154

References

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1-26.

Cleary, T. A, Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist, 30, 15-41.*

Gibson, W.M., & Caplinger, J.A. (2007). Transportation of validation results. In S.M. McPhail (Ed.), *Alternative validation strategies: Developing new and leveraging existing validity evidence.* San Francisco: John Wiley & Sons, Inc.

Hartigan, J. A., & Wigdor, A. K. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the general aptitude test battery. Washington, DC: National Academy Press.*

Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72-98.

Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.

Landy, F.J., Shankster, L.J., & Kohler, S.S. (1994). Personnel selection and placement. *Annual Review of Psychology*, *45*, 261-296.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43, 335-354.*

Ones, D.S., Viswesvaran, C., & Schmidt, F.L. Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, *78*(4), 679-703.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124(2), 262-274.

Schmidt, F.L., Ones, D.O., & Hunter, J.E. (1992). Personnel selection. *Annual Review of Psychology*, *43*, 627-670.