In-Basket Criterion-Related Validity: A Meta-Analysis

Deborah L. Whetzel, Ph.D.

International Personnel Assessment Council
Capitol Ideas for Assessment

July 17-20, 2011

HumRRO
Human Resources Research Organization

Agenda

- Introduction
- How reliable are in-baskets?
- How accurate are in-baskets in predicting performance?
  – Meta-analysis of validity of in-baskets
- What attributes of in-baskets affect validity?
- What attributes of validity studies affect validity?
- To what extent are in-baskets correlated with *g*?
- Are results influenced by publication bias?

Criterion-Related Validity Evidence

- Validity of in-baskets
  - Shippman, Prien, Katz (1990) narrative review
- Validity of work samples
  - Roth, Bobko, and McFarland (2005); $\rho = .33$
  - Hunter and Hunter (1984); $\rho = .54$
- Meta-analysis of the validity of in-baskets
  - Whetzel and Rotenberry (2010)

Validity Evidence

- Conducted a meta-analysis to assess validity of in-baskets
- Literature review
  - Computerized databases (PsycInfo)
  - Listservs (SIOP, PTC/NC, PTC/MW, Academy of Management, IPAC, I/O Practitioners network)

Validity Evidence

- Decision Rules
  - Used job and training performance and salary criteria (not starting salary or personal temperament)
  - Did not include studies that reported only an Overall Assessment Rating (OAR) across all exercises
  - Did not include studies that reported only statistically significant validities
- Number of validity coefficients for each criterion
  - Job performance (k = 32; N = 3,986)
  - Training performance (k = 8; N = 1,563)
  - Salary (k = 14; N = 1,624)

Validity Evidence

- Inter-rater agreement
  - 2 independent coders
  - 190 data points; 18 “disagreements”
  - 90.5% agreement
- Meta-analysis method
  - Hunter and Schmidt (2004)
  - Corrections for criterion unreliability
    - Job performance distribution (Pearlman, Schmidt, & Hunter, 1980); average = .60
    - Training performance distribution (Pearlman, Schmidt, & Hunter, 1980); average = .80
    - Salary was assumed to be perfectly reliable at 1.0

Moderators of validity

- Characteristics of the in-basket
  - Scoring (objective vs. subjective)
  - Content (job-specific vs. generic)
- Characteristics of the study
  - Design (predictive vs. concurrent)
  - Source (published vs. unpublished)

Reliability of In-Baskets

- Two methods for computing reliability
  - Inter-rater reliability (agreement across raters). This is good for methods of multiple constructs
  - Coefficient alpha (internal consistency). This is good for unidimensional measures, such as cognitive ability, or conscientiousness
- In-basket is a method that can measure any number of constructs

Reliability of in-baskets

			Bare Bones Meta-Analysis			
					80% Credibility Interval	
	N	k	$\bar{r}$	SD$\bar{r}$	Lower	Upper
Reliability (interrater)	3,159	28	.77	.15	.58	.97
Reliability (alpha)	2,410	18	.65	.17	.44	.86

Estimated Population Validity of In-Baskets

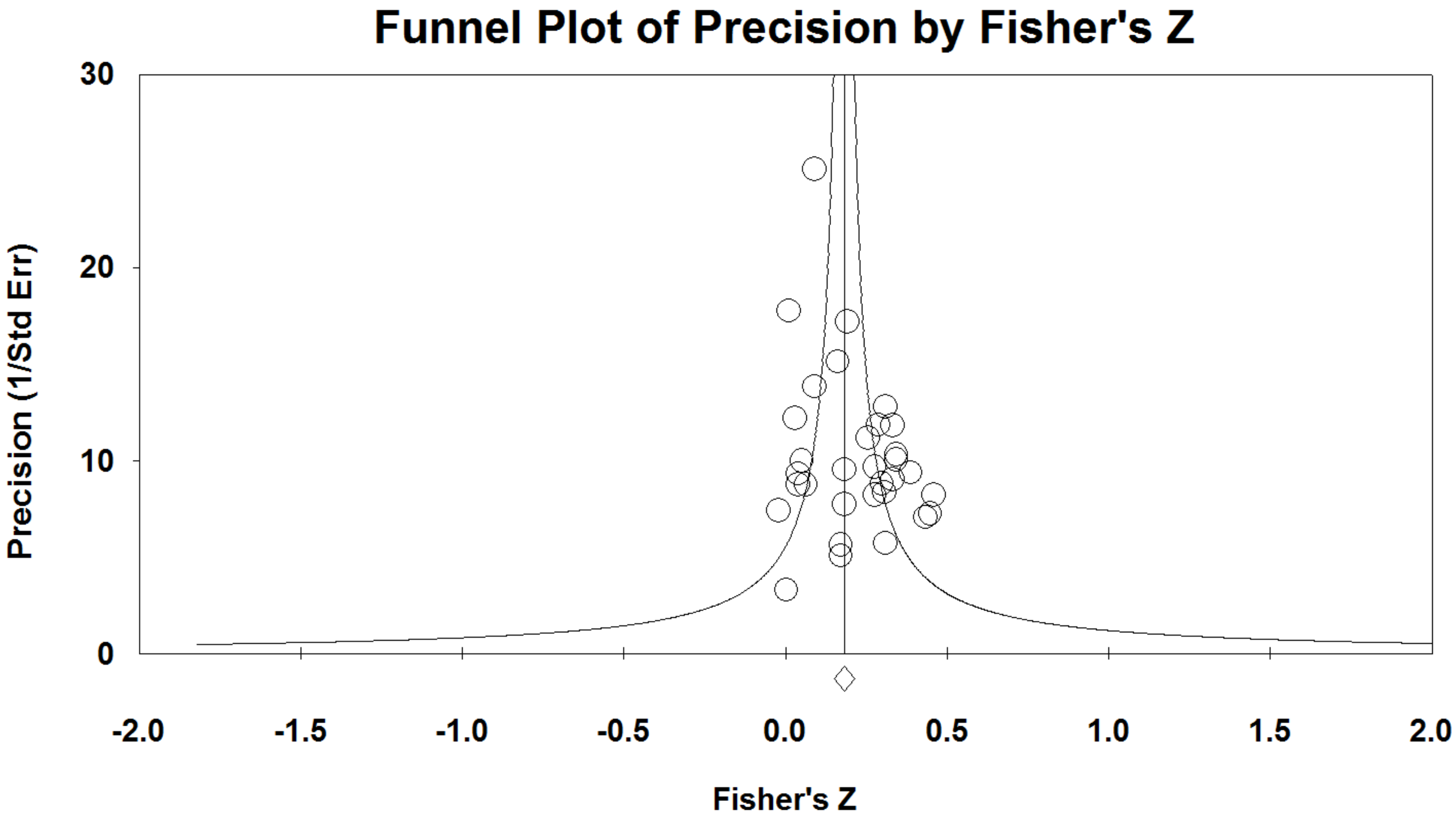
Criterion			Bare Bones Meta-Analysis		Corrected for Criterion Unreliability			
							80% Credibility Interval	
	N	k	$\bar{r}$	SD $\bar{r}$	ρ	SD ρ	Lower	Upper
Job Performance	3,986	32	.18	.09	.36	.14	.19	.54
Training Performance	1,563	8	.17	.05	.31	.06	.22	.39
Salary	1,624	14	.14	.11	.23	.16	.03	.44

HumRRO
Human Resources Research Organization

Publication Bias

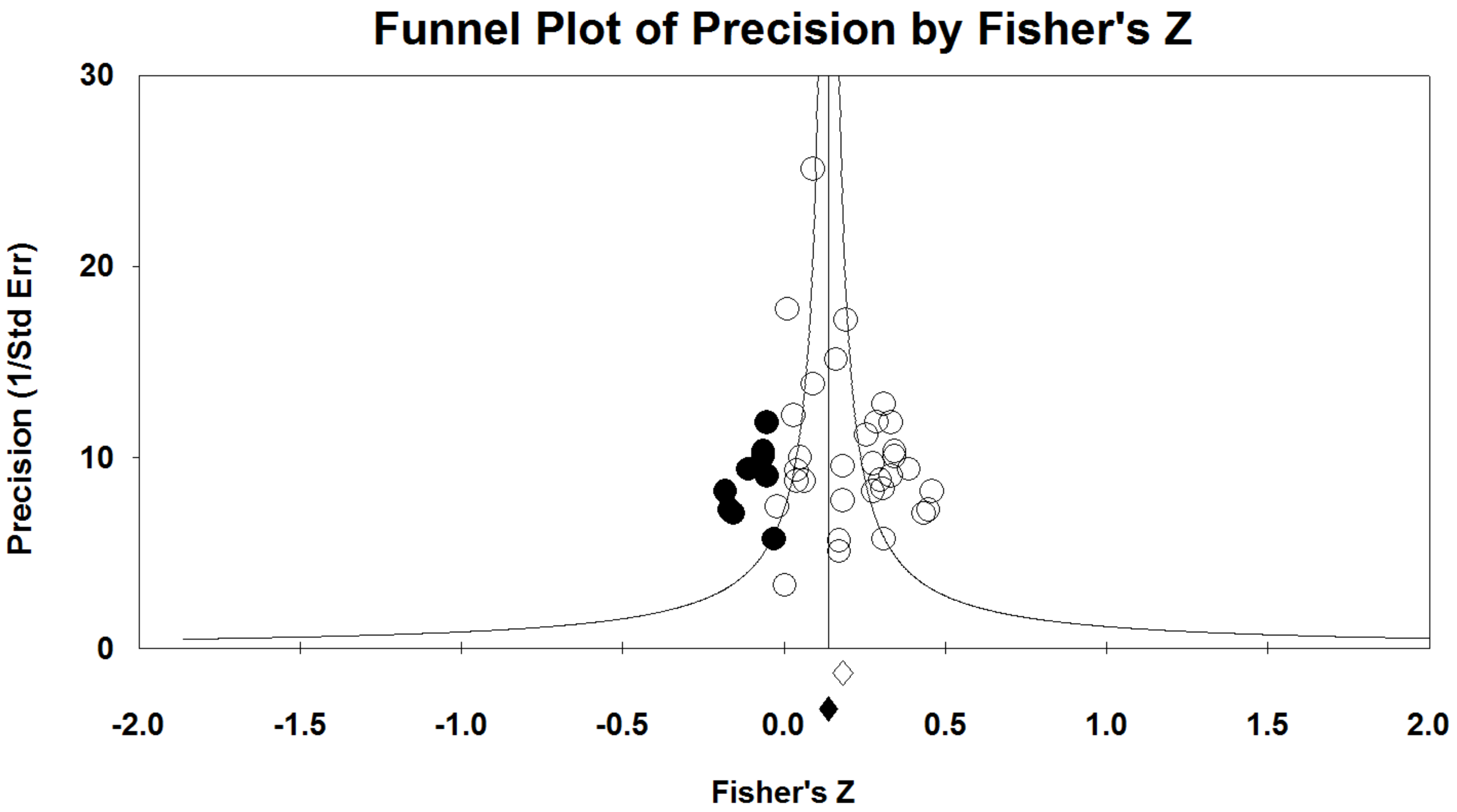
- Exists when the research that appears in the published literature is systematically unrepresentative of the population of completed studies
- The funnel plot
  - X axis displays the magnitude of the effect size
  - Y axis displays precision (highly correlated with sample size)
  - Distribution will be symmetrical if sampling error is only cause of variance

Publication Bias Results



HumRRO Human Resources Research Organization

Publication Bias Results



Job Performance criterion—9 studies imputed; change in validity .06—not much evidence of publication bias

Publication Bias Effects on Observed Validity

Criterion			Publication Bias Analyses Random Effects Model				
	N	k	$\bar{r}$	Studies imputed	Δr	Adjusted observed validity	Higgins I^2
Job Performance	3,986	32	.20	9	.06	.14	51
Training Performance	1,563	8	.18	2	.02	.16	46
Salary	1,624	14	.19	4	.09	.10	63

Estimated Population Validity of In-Baskets

In-Basket Moderator			Bare Bones Meta-Analysis		Corrected for Criterion Unreliability			
							80% Credibility Interval	
	N	k	$\bar{r}$	$SD_{\bar{r}}$	ρ	SD_{ρ}	Lower	Upper
Objective	1,125	12	.15	.09	.31	.15	.11	.51
Subjective	2,230	16	.18	.09	.36	.15	.17	.56
Job-specific	1,916	18	.19	.10	.39	.16	.19	.59
Generic	2,070	14	.16	.07	.34	.11	.21	.48

HumRRO
Human Resources Research Organization

Publication Bias Effects on Observed Validity

In-Basket Moderator			Publication Bias Analyses Random Effects Model				
	N	k	$\bar{r}$	Studies imputed	Δr	Adjusted observed validity	Higgins I^2
Objective	1,125	12	.16	2	.04	.12	50
Subjective	2,230	16	.23	7	.11	.12	60
Job-specific	1,916	18	.22	6	.04	.18	55
Generic	2,070	14	.19	2	.03	.16	49

HumRRO
Human Resources Research Organization

Estimated Population Validity of In-Baskets

Study Feature Moderator			Bare Bones Meta-Analysis		Corrected for Criterion Unreliability			
							80% Credibility Interval	
	N	k	$\bar{r}$	SD $\bar{r}$	ρ	SD ρ	Lower	Upper
Predictive	897	10	.11	.12	.23	.22	-.06	.51
Concurrent	3,089	22	.20	.06	.41	.07	.32	.49
Published	2,547	18	.17	.09	.35	.14	.17	.53
Unpublished	1,974	14	.15	.10	.31	.17	.08	.53

Publication Bias Effects on Observed Validity

Study Feature Moderator			Publication Bias Analyses Random Effects Model				
	N	k	$\bar{r}$	Studies imputed	Δr	Adjusted observed validity	Higgins I^2
Predictive	897	10	.16	1	.04	.12	62
Concurrent	3,089	22	.22	8	.06	.16	40
Published	2,547	18	.21	7	.09	.12	56
Unpublished	1,974	14	.20	0	.00	.20	47

HumRRO
Human Resources Research Organization

Correlation with g

	N	k	Bare Bones Meta-Analysis		Corrected for in-basket unreliability			
			$\bar{r}$	SD $\bar{r}$	ρ	SDρ	Lower	Upper
Correlation with *g*	2,906	18	.26	.08	.30	.09	.19	.41

Effect of Range Restriction

Range Restriction			Bare Bones Meta-Analysis		Corrected for Criterion Unreliability			
							80% Credibility Interval	
	N	k	$\bar{r}$	SD$_{\bar{r}}$	ρ	SDρ	Lower	Upper
Concurrent								
u = .405	3,089	22	.20	.06	.55	.00	.55	.55
u = .500	3,089	22	.20	.06	.47	.03	.42	.51
u = .595	3,089	22	.20	.06	.41	.07	.32	.49
Predictive								
u = .405	897	10	.11	.12	.31	.28	-.05	.66
u = .500	897	10	.11	.12	.26	.25	-.06	.58
u = .595	897	10	.11	.12	.23	.22	-.06	.51

Limitations

- Low k and low N
  - Companies may be concerned about risk of doing a criterion-related validation study
  - Results are often proprietary
  - In-baskets are often part of an assessment center and the data are often reported by dimension/competency

Conclusions

- Objective vs. Subjective—not much difference
- Job-specific vs. Concurrent—minimal difference (but in expected direction)
- Predictive vs. Concurrent—pretty large difference; lots of variance around mean validity for predictive
- Published vs. Unpublished—not much difference for population estimates, but pretty large difference for observed after pub bias (but unexpected direction)
- Correlation with *g*—likely subgroup differences

Utility of the In-Basket

Utility Analysis

- Utility analysis is a method for determining the dollar value of a selection method. It answers the question, "How much money is saved or earned using a valid selection method?"
- The formula for calculating utility (Brogden, 1949; Cronbach & Gleser, 1965) is:

$$U = (T\ N_s\ r_{xy}\ SD_y\ Z_x) - C$$

HumRRO
Human Resources Research Organization

Utility Formula

- $U = (T\ N_s\ r_{xy}\ SD_y\ Z_x) - C$ where:
- U = the dollar value (utility) of the selection procedure
- T = number of years that an employee remains on the job (tenure)
- N_s = the number of people hired each year
- r_{xy} = the correlation between the assessment and job performance; the validity of the assessment
- SD_y = the difference between high and low levels of job performance (Research shows 40% of salary)
- Z_x = the score of people above the “cutoff”; ratio of the number of selected applicants to total applicants
- C = cost of developing, validating, and administering the assessment to applicants

HumRRO
Human Resources Research Organization

Utility Example: HR Manager

- T = 10 years (assume HR Manager tenure in an organization is about 10 years)
- N_s = 2 (assume the average number of HR Managers hired per year in an organization)
- r_{xy} = .23 (predictive validity of in-baskets)
- SD_y = 36,000 (assume the average salary for HR Managers is $90,000; underestimate not including benefits)
- Z_x = .80 (mean of 0 and SD of 1).
- C = Development/Validation Study and administration costs = $10,000.

Utility Results and Implications

- The value to an organization of using an in-basket for the first year is $122,480, assuming
  - 2 HR Managers are hired each year
  - Each one stays for 10 years
  - They make an average of ~$90,000 per year (median salary; O*NET, 2009)
- The difference between good and bad HR Managers is about 40% of their annual salary.
- While a savings of $122,480 may seem high, think of the critical hire/fire decisions an HR manager makes and the advice they provide regarding legal HR issues

References

Brogden, H.E. (1949). When testing pays off. *Personnel Psychology, 2*, 171-183.

Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana, IL: University of Illinois Press.

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

Hunter, J.E. & Schmidt, F.L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.

Roth, PL., Bobko, P., McFarland, L.A. (2005). A meta-analysis of work sample test validity, Updating and integrating some classic literature, *Personnel Psychology, 58*, 1009-1037.

Whetzel, D.L., & Rotenberry, P.F. (2010). *In-basket criterion-related validity: A meta-analysis*. Paper to be presented at the 2011 Annual Convention of the Society of Industrial and Organizational Psychology, Inc., Chicago, IL.

Questions?

Thank you!