

In-baskets: What are they and why should we care?

International Personnel Management Association Conference
Las Vegas, NV

Deborah L. Whetzel, HumRRO
Paul F. Rotenberry, West Chester University
July 23, 2012

HumRRO

Human Resources Research Organization

Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 700 Alexandria, Virginia 22314-1591
Phone: 703.549.3611 | Fax: 703.549.9661 | www.humrro.org

Introduction

- Human Resources Research Organization (HumRRO)
 - Non-profit research organization
 - Over 60 years of experience
 - About 100+ employees (I/O psychologists, psychometricians, educational psychologists)
 - Developed assessment centers and in-baskets for
 - Bureau of Alcohol, Tobacco, Firearms and Explosives
 - Social Security Administration
 - Federal Bureau of Investigation



Introduction

- **What are in-baskets?**
 - Relatively high fidelity simulations
 - Used to predict performance in many occupations including law enforcement, managerial, clerical and variety of professional jobs
 - Widely used due to face validity, ease of administration and availability online
 - Often used as part of assessment centers

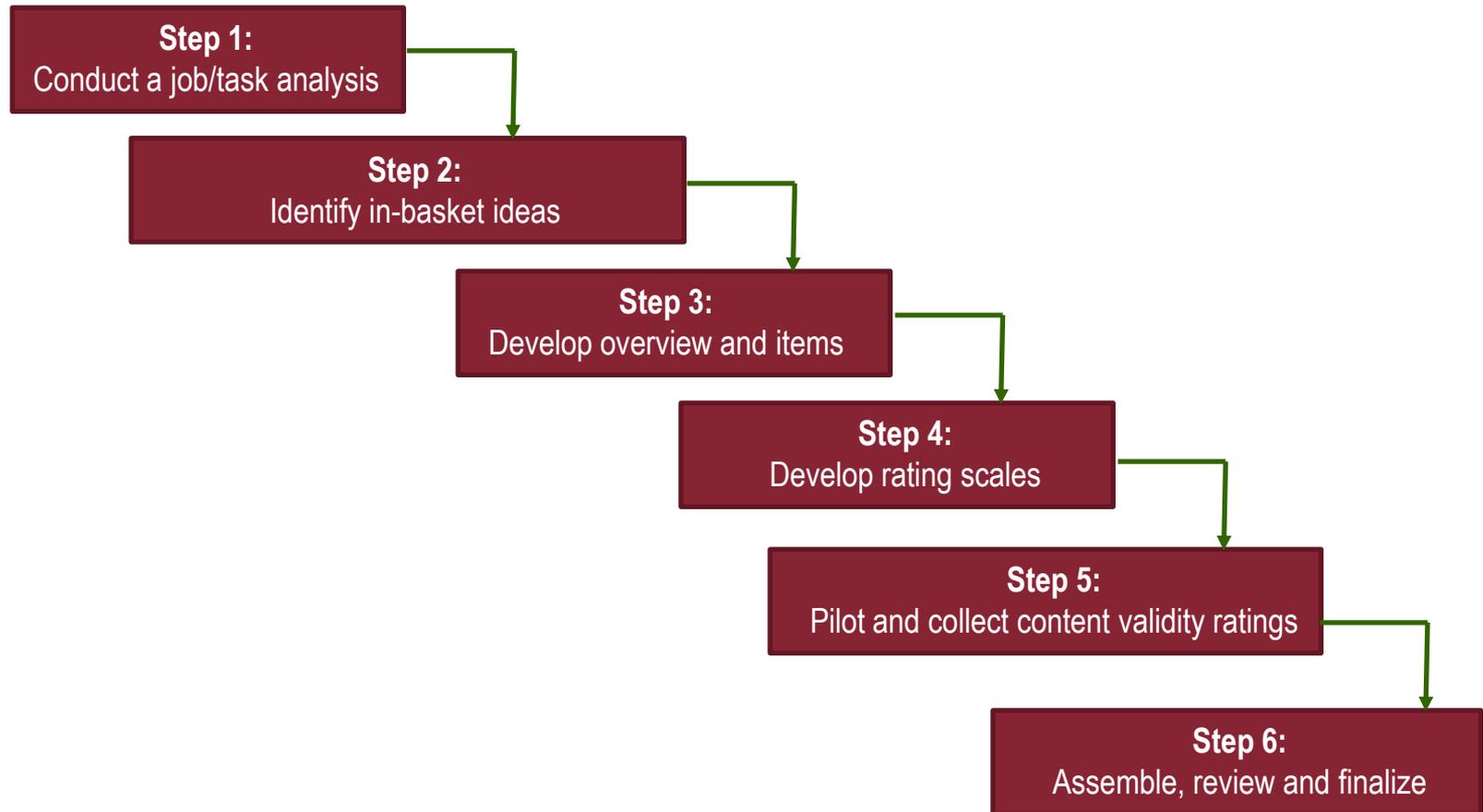


Content Validity Strategy

Evidence for validity based on content typically consists of a demonstration of a strong linkage between the content of the selection procedure and important work behaviors, activities, worker requirements, or outcomes on the job.

--SIOP Principles, p. 21

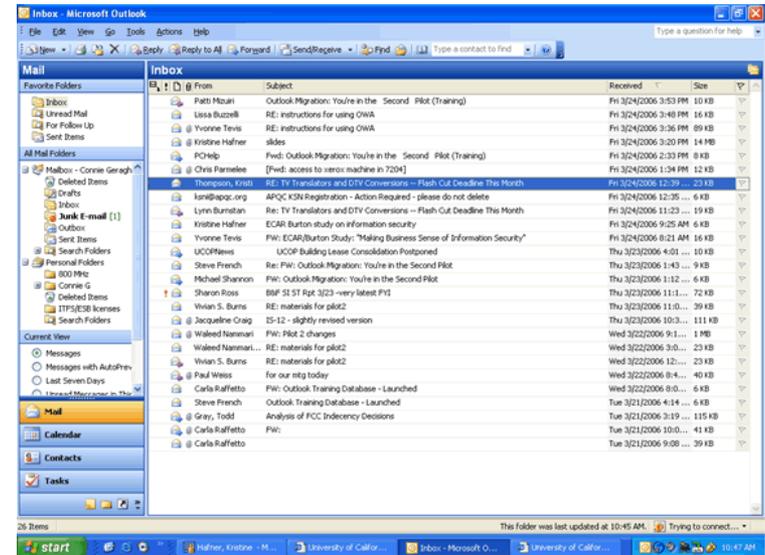
Steps to Developing Content Valid In-Baskets



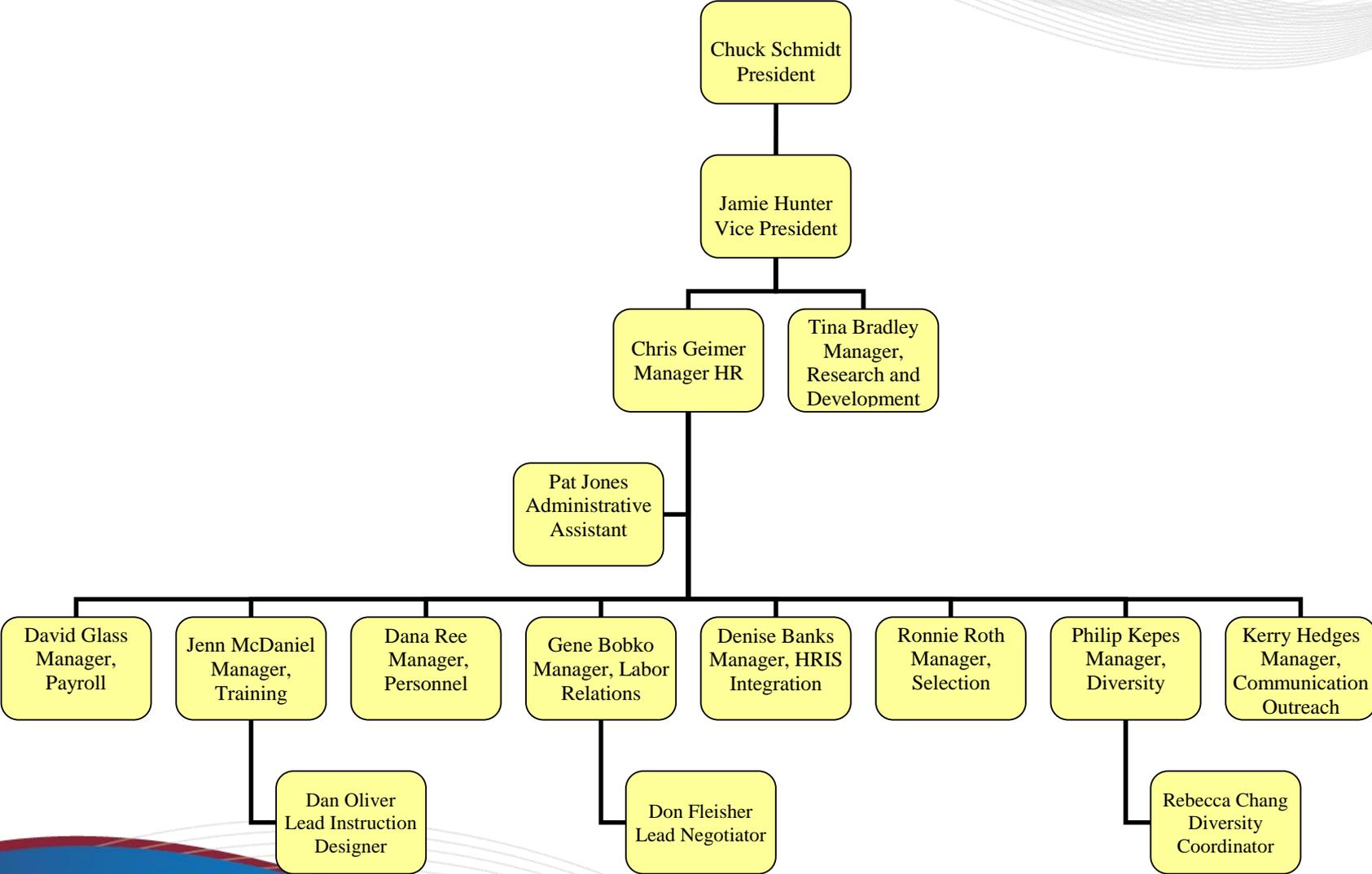
Introduction

■ How do in-baskets work?

- Examinees are given documents typically found in an in-basket or in-box
- Examinees respond to each document by indicating the actions they would take
- Responses include prioritizing tasks, identifying inconsistencies, finding mistakes, delegating, handling problems



Example for HR Manager of Feuquay Enterprises



Example for HR Manager of Feuquay Enterprises

Feuquay Enterprises		
HR Management Personnel	Information	Tenure with Feuquay and in current position
David Glass, Manager, Payroll	<ul style="list-style-type: none"> ▪ Been around for a long time ▪ May be “retired in place” 	<ul style="list-style-type: none"> ▪ 13 years at Feuquay ▪ 10 years in current position
Jenn McDaniel, Manager, Training	<ul style="list-style-type: none"> ▪ Very conscientious ▪ May drive employees too hard 	<ul style="list-style-type: none"> ▪ 7 years at Feuquay ▪ 2 years in current position
Dana Ree, Manager, Personnel	<ul style="list-style-type: none"> ▪ Has worked in personnel 10 years ▪ Instituted many procedures still used 	<ul style="list-style-type: none"> ▪ 15 years in Feuquay ▪ 10 years in current position
Gene Bobko, Manager, Labor Relations	<ul style="list-style-type: none"> ▪ Recently promoted to management ▪ Used to be a union representative 	<ul style="list-style-type: none"> ▪ 10 years at Feuquay ▪ 3 months in current position
Ronnie Roth, Manager, Selection	<ul style="list-style-type: none"> ▪ New to his current position. ▪ Was a Lead Psychologist in charge of developing selection systems 	<ul style="list-style-type: none"> ▪ 5 years at Feuquay ▪ New to current position
Pat Jones, Admin Assistant	<ul style="list-style-type: none"> ▪ Lots of experience in Feuquay ▪ Conscientious and hard working 	<ul style="list-style-type: none"> ▪ 9 years with Feuquay ▪ 6 years in current position

E-mail item #3 from Manager, Selection

Hi Chris,

As you have known for some time, we need to do a full-scale test development and validation effort for the position of Software Programmer. There have been several complaints that new employees do not have skills needed to do the job. Currently, we use interviews to select people for this job.

A new selection system will require a job analysis to determine what Software Programmers do and to identify knowledges, skills, and abilities (KSAs) needed to do the job. We will then need to develop (or purchase off-the-shelf) tests that assess the KSAs identified. If we decide to conduct a criterion-related validation study, we will need to get employees to take the tests and get their supervisors to provide performance ratings (for research purposes only, of course).

To accomplish this, we will probably need to hire at least one additional Industrial Psychologist and several of our current employees will need to work on this full time.

Thanks, Ronnie Roth

Voicemail item #8 from Admin, Pat Jones

Hi Chris, It's Pat Jones. Congratulations on your promotion! They definitely picked the right person for the job. You've been a terrific mentor so far and I'm hoping to learn a lot from you in your new position. As you know, I just got my Masters' degree in Labor Relations and did an internship with CFG International. How about lunch later this week?

E-mail item #13 from Manager, Labor Relations

Chris,

As you know, we negotiate with the labor unions when their contracts expire, every three years. Our bargaining sessions will start this April and because of the decline in business, I'm expecting this to be a fairly contentious negotiation. It will take my entire staff most of this month to prepare for these sessions.

At the same time, we have several reports that are due to Diversity at the end of March regarding the composition of the workforce. Finally, we need to respond to the problem at the Langan plant regarding the low scores on the employee satisfaction survey. We also have lost a couple of key staff members who have a lot of experience with the unions and I will need to hire people from outside Feuquay who have labor relations experience. Please advise on how to manage this workload.

Thanks, Gene Bobko

Example rating scale for Resources Management

Resources Management

Ability to prioritize resources (e.g., personnel, funds); identifies need for additions/reductions in staffing levels

1	2	3	4	5
LOW	MODERATE		HIGH	
<ul style="list-style-type: none"> ➤ Accepted Ronnie Roth's request without question (#3) ➤ Failed to notice connection with #8 and possible opportunity for Pat Jones (#13) 	<ul style="list-style-type: none"> ➤ Asked Ronnie Roth for more information about the nature of the complaints (#3) ➤ Noticed that Kerry needs help in Labor Relations and made connection regarding Pat's newly earned Masters' degree (#13) 		<ul style="list-style-type: none"> ➤ Probed Ronnie Roth about complaints and determined whether selection was the issue (#3) ➤ Called Kerry and suggested that he consider Pat Jones for a job in Labor Relations and made it clear that if Pat didn't have sufficient experience, he was authorized to go outside of Feuquay to hire (#13). 	

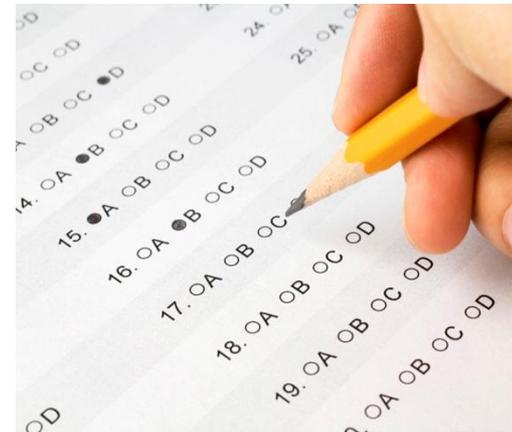
Introduction

- **What is the purpose of this study?**
 - To determine the validity of in-baskets for predicting job performance
- **What makes this study awesome and cool?**
 - Compiled largest known database (32 studies; 3,986 people)
 - Tested validity using four moderators
 - Conducted sensitivity analyses to determine how robust results were to publication bias

Moderators

- **Objective vs. subjective**

- Objective = minimal use of human judgment; checklists
- Subjective = some use of human judgment; rating scales
- Hypothesis 1: In-baskets that are subjectively scored will yield higher validity estimates than in-baskets that are objectively scored.



Moderators

■ Job Specific vs. Generic Content

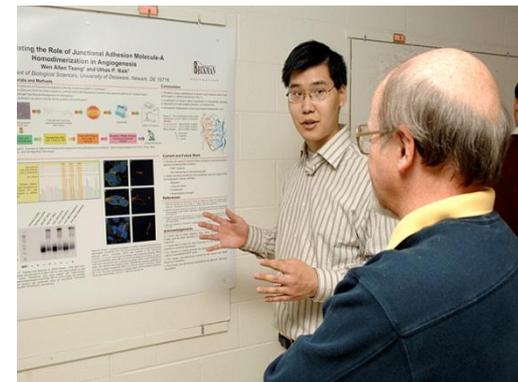
- Job-specific = in-baskets designed for a specific job using job analysis
- Generic = in-baskets designed for multiple jobs (e.g. management jobs)
- Hypothesis 2: In-baskets with job-specific content will yield higher validities than in-baskets with generic content.

Moderators

■ Published vs. Unpublished

- Published = journal articles and book chapters
- Unpublished = conference presentations and technical reports
- Hypothesis 3: Validity studies that have been published will yield higher validities than studies that have not been published.

Publications



Moderators

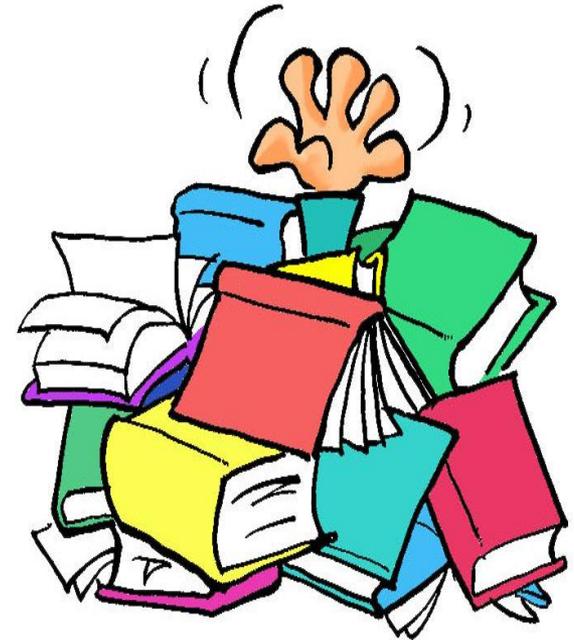
■ Concurrent vs. Predictive

- Concurrent = administer predictor and criterion measures at same time
- Predictive = administer predictor and concurrent at 2 points in time
- Hypothesis 4: Studies using a concurrent design will yield higher validities than studies using a predictive design.

Concurrent	Predictive
Collect predictor and criterion data simultaneously	Time lag between collecting predictor and criterion data
Collect data on job incumbents	Collect data on job applicants
Range restriction	No range restriction

Method

- **Conducted literature search**
 - Keyword search using PsycINFO
 - Reference lists from in-basket studies
 - Calls for papers from listserves
 - SIOP
 - **IPAC**
 - I/O Practitioners Network
 - I/O Careers
 - PTC/MW, PTC/NC, PTC/SC
 - NY Metro
 - Chicago I/O Psychologists
 - Academy of Management (AOM)



■ Decision Rules

- Did not include studies that reported only an Overall Assessment Rating (OAR)
- Did not include studies that only reported statistically significant validity coefficients
- Did not include studies that used temperament, interests, and start salary as criteria



Method

■ Interrater Agreement

- Two researchers independently coded all studies for N, r, and criterion
- Out of 190 data points, there were 18 “disagreements” resulting in a .91 level of agreement
- Discussed and resolved by referring to or modifying decision rules



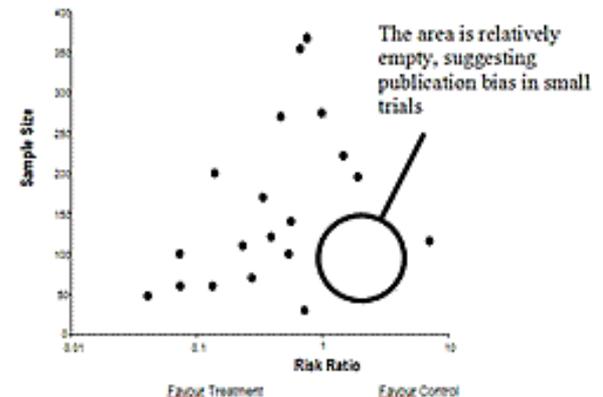
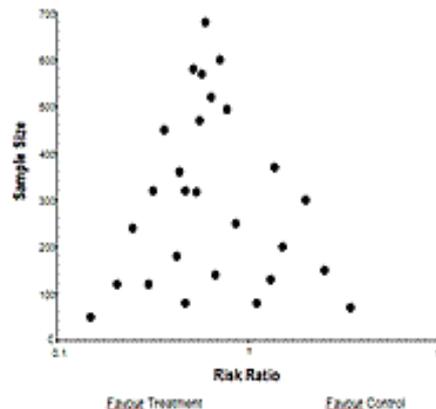
Method

- **Meta-analysis Techniques**

- Schmidt and Le (2005) program; corrected for criterion unreliability using Pearlman's (1980) assumed distribution

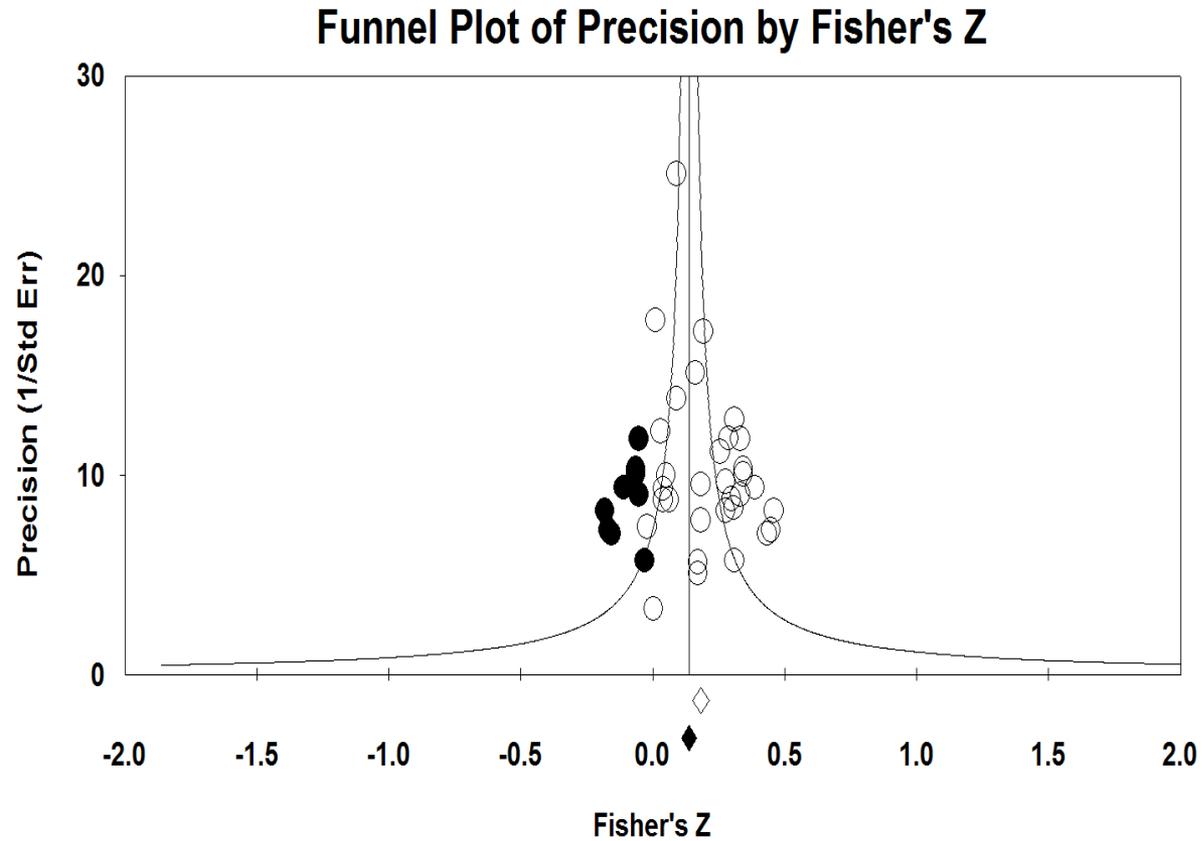
- **Publication Bias**

- Comprehensive meta-analysis (Bornstein, Hedges, Higgins, & Rothstein, 2005); trim and fill (Duval & Tweedie, 2000a,b)



- **Publication Bias using Trim and Fill**
 - Evaluates the degree of symmetry in a funnel plot of validities
 - “Trims” extreme validities from the skewed side of the sampling distribution in the funnel plot
 - “Fills” in the trimmed validities on the opposite side needed to achieve symmetry
 - Re-estimates validity in the potential absence of publication bias

- Publication Bias



Results—Job Performance Criterion

	N	k	r	SDr	Population estimates		80% credibility interval		Trim and Fill			
					ρ	$SD\rho$	Lower	Upper	REr	# studies imputed	Δr	T&F adjusted
Job Perf (all studies)	3,986	32	.18	.09	.36	.13	.19	.53	.20	3,986	32	.18
Job Perf (no outlier)	3,353	31	.19	.09	.40	.12	.24	.55	.21	3	.02	.19

- Level of validity is .18 uncorrected; .36 corrected

Results—Moderated by Scoring Method

Scoring Method	N	k	r	SDr	Population estimates		80% credibility interval		Trim and Fill			
					ρ	$SD\rho$	Lower	Upper	REr	# studies imputed	Δr	T&F adjusted
Objective	1,125	12	.15	.09	.31	.15	.11	.51	.16	2	.04	.12
Subjective	2,230	16	.18	.09	.36	.15	.17	.56	.23	7	.11	.12

- No moderator effect of in-basket **scoring** (objective vs. subjective) after adjusting validities for publication bias.

Results—Moderated by Content

In-basket Content	N	k	r	SDr	Population estimates		80% credibility interval		Trim and Fill			
					ρ	$SD\rho$	Lower	Upper	REr	# studies imputed	Δr	T&F adjusted
Job-Specific	1,916	18	.19	.10	.39	.16	.19	.59	.22	6	.04	.18
Generic	2,070	14	.16	.07	.34	.11	.21	.48	.19	2	.03	.16

- Minimal moderator effect for **content**, although the validities fell in the expected direction (i.e., job-specific in-baskets yielded higher validity estimates than generic in-baskets). This result was consistent after adjusting for publication bias.

Results—Moderated by Data Source

Data Source	N	k	r	SDr	Population estimates		80% credibility interval		Trim and Fill			
					ρ	$SD\rho$	Lower	Upper	REr	# studies imputed	Δr	T&F adjusted
Published	2,547	18	.17	.09	.35	.14	.17	.53	.21	7	.09	.18
Unpublished	1,439	14	.19	.14	.39	.12	.24	.55	.20	0	.00	.16

- **Data source** did act as a moderator; unpublished studies yielded higher validity coefficients than published studies. This was true after accounting for publication bias.

Results—Moderated by Study Design

Study Design	N	k	r	SDr	Population estimates		80% credibility interval		Trim and Fill			
					ρ	$SD\rho$	Lower	Upper	REr	# studies imputed	Δr	T&F adjusted
Predictive	897	10	.11	.12	.23	.22	-.06	.51	.16	1	.04	.12
Concurrent	3,089	22	.20	.06	.41	.07	.32	.49	.22	8	.06	.16

- Study design.** Concurrent studies yielded higher validity estimates than predictive studies. Likely due to predictor contamination (those who score in-basket exercises in concurrent studies may artificially adjust a person's score based on their knowledge of the person's job performance, thus contributing to higher levels of validity for concurrent studies).

Limitations

- Low k and low N
 - Companies may be concerned about risk of doing a criterion-related validation study
 - Results are often proprietary
- Validities are underestimates due to lack of range restriction (direct and indirect) corrections

Utility of the In-Basket

- 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:

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

Utility of the In-Basket

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

Utility Example: HR Manager

- $T = 10$ years (assume HR Manager tenure in an organization is 10 years)
- $N_s = 2$ (assume the average number of HR Managers hired per year in an organization)
- $r_{xy} = .36$ (the corrected correlation between the assessment and job performance; the validity of the assessment)
- $SD_y = 36,000$ (assume the average salary for HR Managers is \$90,000, underestimate not including benefits)
- $Z_x = .80$ (assume a selection ratio of .50—50% of the people who apply for an HR manager job are selected; Z is the average of people above the cutoff core (the top half of the distribution))
- $C = \$10,000$ cost of developing, validating and administering in-basket

Utility Results and Implications

- The value to an organization of using an in-basket over random selection is \$197,360 assuming:
 - 2 HR Managers are hired each year
 - Each one stays with Feuquay 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 \$197,000 may seem high, think of the critical hire/fire decisions an HR manager makes and the advice they provide regarding legal HR issues.

References

- Borenstein M., Hedges L., Higgins J., & Rothstein, H.R. (2005). *Comprehensive meta-analysis. Version 2*. Englewood, NJ: Biostat.
- 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.
- Duval S.J., & Tweedie RL. (2000a). A non-parametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89–98.
- Duval S.J., & Tweedie RL. (2000b). Trim and fill: A simple funnel plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 276–284.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance, *Psychological Bulletin*, 96, 72-98.

References

- Hunter, J.E. & Schmidt, F.L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.
- Pearlman, K., Schmidt, F.L., & Hunter, J.E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65, 373-406.
- Roth, P.L., 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.
- Schmidt, F.L., & Le, H. (2005, October). Hunter and Schmidt meta-analysis programs (V 1.1)
- Whetzel, D.L., & Rotenberry, P.F. (2012). *Moderators of in-basket criterion-related validity: A meta-analysis*. Paper presented at the 2012 Annual Convention of the Society of Industrial and Organizational Psychology, Inc., Chicago, IL.



Thanks!