

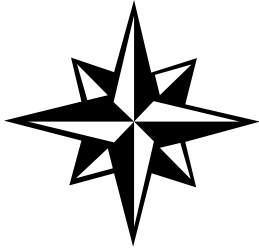
Capturing Performance: Developing Criterion Measures for Validation Studies

Julia McElreath

U.S. Department of Homeland Security

Theodore L. Hayes

Transportation Security Administration



Overview of Presentation

- **General discussion of criteria**
- **How we approached the criterion issue**
- **Development of three criterion measures**
- **Results of validation studies**
- **Lessons learned**

What are Criterion Measures?

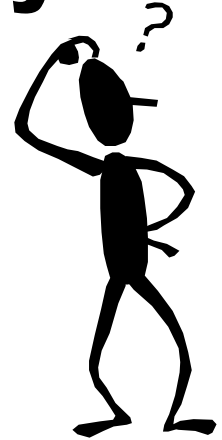
- **Indicators of job performance**
- **According to the Uniform Guidelines, criterion measures in validation research must:**
 - **represent critical job duties or outcomes**
 - **not be biased**
- **One tool used to determine appropriate selection batteries**

Examples of Criteria

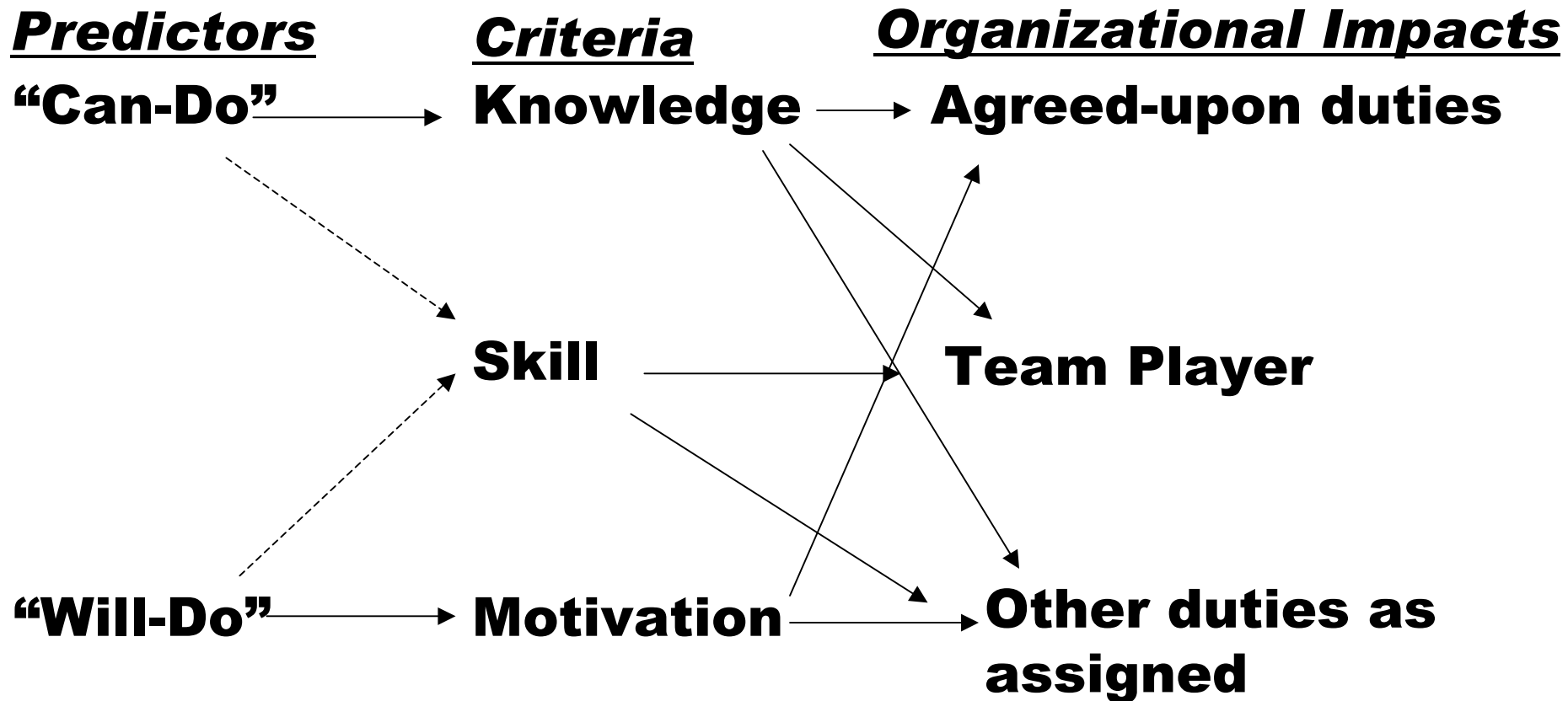
- **Performance Ratings**
- **Training Scores**
- **‘Job Performance’ Scores**
- **Productivity**
- **Absences**
- **Errors**
- **Customer Satisfaction**

The Criterion Problem(s)

- **What do the criteria actually measure?**
- **How meaningful are the criteria?**
- **Should there be one (unidimensional) criterion or multiple criterion measures?**



A Model of Job Performance*



*After Schmitt, Cortina, Ingerick, & Wiechmann, 2003

Our Approach to Criteria

- **Validation of selection batteries for multiple occupations in a law enforcement agency**
- **Each study began with a foundation in job analysis**
- **Performance is multidimensional in each occupation - we needed multiple criteria**
 - **Training Scores**
 - **Work Simulation Scores**
 - **Supervisory Performance Ratings**

Training Scores

- **Training for each occupation is extensive**
- **Focus on learning relevant law**
- **Content of courses changes as laws change, but focus and structure of courses are constant**
- **Tests are given regularly in training courses; final grades are reliable, objective, and consistently applied**
- **Average of two law classes**

Work Simulation Exercise

- **Multiple choice questions containing:**
 - **Scenarios incumbents are likely to face on the job**
 - **Five response options for each scenario**
- **Subject Matter Experts develop scenarios and response options**
- **Multiple reviews ensure accuracy and relevance of questions and answers**
- **SMEs are representative of work force**

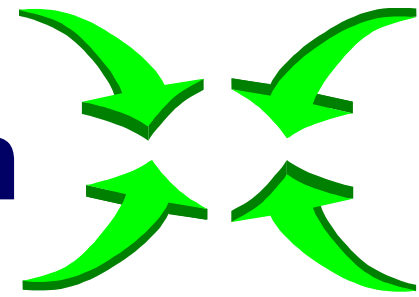
Supervisory Performance Ratings

- **Developed by Subject Matter Experts**
- **Written for competencies and duties evaluated as critical for successful performance**
- **Behaviorally anchored 15-point scale; anchors for unsatisfactory, satisfactory, and excellent performance**
- **Written for level of validation study participants**
- **Overall performance = average of all performance dimensions**

Linking The Model With Our Criteria

- **Knowledge: knowledge of facts and procedures**
 - **Training, Work Simulation**
- **Skill: knowledge of how to do something as well as what to do**
 - **Work Simulation, Supervisor Ratings**
- **Motivation: includes knowing what to do, how much effort to put forth, and for how long to work on a task**
 - **Supervisor Ratings**

Composite Criterion



- **Sum of T-Scores for training, work simulation, and average performance ratings**
- **Composite criteria should not be used alone...**
 - **Composite criteria make sense after correlations among criteria are evaluated**
 - **Composite criteria do not make sense when criterion measures do not correlate well - combining criteria dulls predictions (Guion, 1998)**
 - **Using ONLY composite criteria may mask predictive ability of individual tests**

Validation Studies

- **Four occupations**
- **Multiple predictors and criterion measures in each study**
- **Predictors:**
 - **Cognitive and biodata**
- **Criterion Measures:**
 - **Training scores, work simulation, performance ratings, composite criterion**

Correlations Among Criterion Measures

	training	work sim	ratings
training	---	.36	.23
work sim		---	.18
ratings			---

Note: These are weighted averages of obtained correlations from the four validation samples

Results for Training Score Criterion

	With cognitive	With biodata
● Occ. A	.58	.25
● Occ. B	.50	.15
● Occ. C	.64	.38
● Occ. D	.66	.23

Results for Work Simulation Criterion

	With cognitive	With biodata
● Occ. A	.48	.20
● Occ. B	.49	.17
● Occ. C	.62	.24
● Occ. D	.57	.12

Results for Performance Ratings Criterion

	With cognitive	With biodata
● Occ. A	.28	.20
● Occ. B	.04	.26
● Occ. C	.37	.37
● Occ. D	.25	.30

Results for Composite Criterion

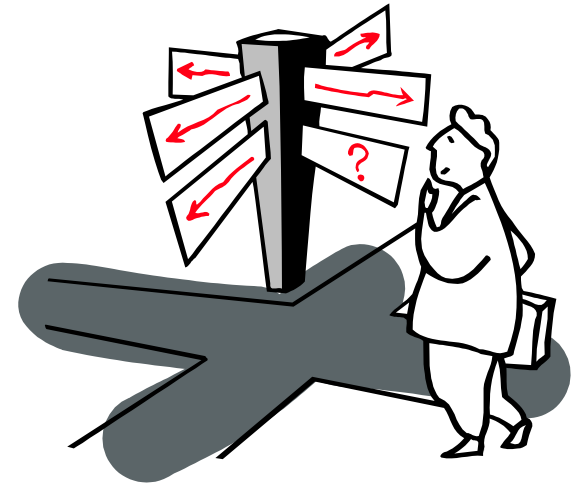
	With cognitive	With biodata
● Occ. A	.64	.30
● Occ. B	.66	.39
● Occ. C	.51	.26
● Occ. D	.63	.26

Overall Pattern of Results

- **Training and work simulation criteria more related to cognitive tests than to biodata measures**
- **Supervisory ratings generally more related to biodata than to cognitive predictors**
- **Higher composite criterion validities for cognitive measures than for biodata**

Use of Criterion-Related Validity Evidence

- **One factor in determining selection battery**
- **Other factors:**
 - **reliability of tests**
 - **fairness**
 - **pass rates**
 - **impact of cut scores on organizational outcomes**



Usefulness of Multiple Criteria

- **Allowed for measurement of performance from more than one perspective**
 - **Joe: work simulation score = 5 (out of 40), training score = 70% (minimum passing), and performance ratings = 14.8 (out of 15)**
 - **Q: is Joe a desirable employee?**
- **Allowed for evaluation of relationships between different predictors and different aspects of performance**
- **Permitted assembly of test batteries that are indicative of well-rounded applicants**

Lessons Learned



- **Developing multiple criteria is critical for maximizing opportunities for identifying optimal test batteries in validation studies**
- **Multiple criteria can be developed relatively quickly and with minimal intrusion on incumbents**
- **Multiple criteria are credible, make sense to decision makers, and reflect good science practice**
- **Patterns of relationships hold across samples; criteria are relatively robust**