Employment Test Validation
Introduction

- Yari Randall, Test Development Manager
  - M.S., Applied Industrial and Organizational Psychology, University of Baltimore
  - B.S., Psychology, University of Maryland at College Park
  - Global Organizational Structure Development Certification, Barcelona, Spain
What are employment assessments?

- Objective, standardized way of assessing candidates’ qualifications and/or predicting candidate performance on-the-job.
- Entry-level or Promotional
Types of Employment Tests

- **General Cognitive Aptitude Tests**
  - Critical Thinking, Problem-solving, Ability to Learn and Apply Information
- **Personality Tests**
  - Big 5 (OCEAN): Openness to Experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism; Integrity Tests
- **Skill Tests**
  - Job specific competencies such as verbal, math or accounting, or data-entry skills
Resume checks and interviews are often inefficient and unreliable.

Increase productivity by successfully predicting performance.

Reduce costs associated with turnover (e.g. hiring and training costs).

Streamline hiring process.

Increase legal defensibility of your hiring process.

Why Use Employment Tests?
Let’s Talk About Legal Defensibility

Title VII of the Civil Rights Act of 1964
Important Legal Precedents

- Albemarle Paper Co. v. Moody (1975)
- Washington v. Davis (1976)
- Ricci v. Destefano (2009)
Important Legal and Professional Guidelines

  - Adopted Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures (1979) – EEOC, OPM, DoJ, DoL, and DoT

- Standards for Educational and Psychological Testing (2014) – AERA, APA, & NCME

- Principles for the Validation and Use of Personnel Selection Procedures (2019) - SIOP
Test Validity

- Unitary Concept: “the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed use” (Standards, 2014, p. 13-14).
  - Principles (2019): “a unitary concept with different sources of evidence contributing to an understanding of the inferences that can be drawn from a selection procedure”
- There are no distinct types of validity; rather, there are only sources of evidence that can be used to support validity in general.
Traditional Validity Theory

- Criterion Validity
- Content Validity
- Construct Validity
Modern Validity Theory

• Sources of validity evidence (*Standards*, 2014, p. 14-21)
  • Evidence based on test content
  • Evidence based on response processes
  • Evidence based on internal structure
  • Evidence based on relations to other variables
Validation

- Process of gathering evidence related to the intended interpretation of test scores
- Validation is an ongoing, continuous process: “as validation proceeds, and new evidence regarding the interpretations that can and cannot be drawn from test scores becomes available, revisions may be needed in the test, in the conceptual framework that shapes it, and even in the construct underlying the test” (Standards, 2014, p. 12).
Criterion-related Validation

• Process of gathering evidence based on the test’s relationship to performance criteria (e.g., job performance)

• Previously regarded as the highest source of validity evidence (Sproule, 2009)
  • “Thirty years or so ago, there was a widespread belief that the best evidence of the validity of a test in a given organization was a ‘local’ [criterion-related] validation study conducted in that organization for that job. But the number of employees available in any particular organization on any particular job is almost always so small that the results of the local [criterion-related] study are unstable statistically” (Schmidt, 2012).
<table>
<thead>
<tr>
<th>Concurrent Criterion Validation</th>
<th>Predictive Criterion Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Uses current employees</td>
<td>• Uses candidates</td>
</tr>
</tbody>
</table>

Types of Criterion Validation
Limitation #1 of Criterion-related Validation

- **Sample Size**
  - “Sample size affects the degree to which different lines of evidence can be drawn on in examining validity for the intended inference to be drawn on the test. For example, relying on the local setting for empirical linkages between test and criterion scores is not technically feasible with small sample sizes” *(Standards, 2014, p. 171).*

*Important: Relying on a criterion-related validation study with too small of a sample size as your only piece of validation evidence is not likely to be viewed as an acceptable professional practice.*
- So how big of a sample do I need?
  - Minimum 150 (Biddle & Bell-Pilchard, 2013, p. 14)
  - 2,000 (Schmidt, 2012)
  - “If the number of workers available for the study is less than about 2,000, the estimate of predictive validity will suffer from large statistical sampling error, and as a result, the confidence interval around the validity estimate will be wide.”
  - “In the Russell et al. (1994) meta-analysis of all the criterion-related validity studies published in the Journal of Applied Psychology and Personnel Psychology between 1964 and 1992, the median sample size was 103...A sample size of 103 contains large amounts of sampling error and does not produce statistically stable validity estimates (with N = 103, the width of the 95% confidence interval is approximately .43)."
Limitation #2 of Criterion Validation

• Inadequate Criterion Measures
  • Objective v. Subjective Performance Measures
  • Supervisory Performance Ratings are the most commonly used criterion-related measure
    • What about rater biases? (i.e., leniency, severity, halo effect)
    • What about the politics of the supervisor-subordinate relationship?
    • “Most reviews of performance appraisal research have shown that the relationship between performance ratings and actual job performance is weak or at best uncertain” (Murphy, 2008 as cited in Sproule, 2009)
• **Range restriction**

  “To illustrate the problem of range restriction, most merit hiring is done in score order. Only those with high scores are typically hired. When only a few hires occur, this presents both a sample size problem and a severe range restriction problem. Criterion related validation is not feasible when the sample size is small and the predictor score representation is overly restricted. The scores of those hired do not represent the scores of the applicant group, so we have no way to determine if those with low or moderate scores would be good or poor job performers” (Sproule, 2009).
Limitation #4 of Criterion Validation

• Test Score Interpretations
• “A direct validation study gathers data on job performance to estimate how well it is predicted by various potential methods of selecting applicants. While such studies can demonstrate criterion validity, they do not usually provide an understanding of why the predictor is valid.”
• Revisiting Griggs v. Duke Power
So What Does This Mean?

• Do not rely strictly on your local criterion-validation as your **sole source** of validity evidence, especially if you suffer from the following limitations:
  • Small sample size
  • Inadequate Criterion Measures
  • Range restriction
Content Validation

- Process of establishing validity by studying the psychological tasks and processes underlying work performance, which are often specific cognitive processes and skills (e.g., simple mental arithmetic or mental rotation of objects)

- Has been regarded as the most powerful source of validity evidence from an explanatory point of view because it explains why certain attributes are relevant for a particular job instead of just presenting numbers that show that a test predicts performance (Schmidt, 2012)

- The Standards (2014) recognize content validation as the “primary source of validity evidence” (Standard 11.3, p. 178)
IPAC Views on Content Validity

“It is a validation process that can address the constraints faced by many organizations. It is a practical approach to validation. It prescribes an approach to developing the assessment based on a study of the job. Validity is thus built into the assessment procedures. Assessment methods based on this strategy are usually statistically related to job performance.”

“Content validity is essential to the basic intrinsic meaning of any measure. The criterion measures needed for criterion-related validity must themselves possess content validity. Thus, content validity is a prerequisite for validation.”
How do I Establish Content Validity?

• Job Analysis!
  • You’re doing it anyway
  Subject Matter Expert Reviews of Test
  (relevance; representativeness)
Reservations about Content Validation

• Is it seen as a less professional practice compared to criterion-related validation?
• Will it be as legally defensible in Court?
• Is content validation appropriate for the type of test I am developing?
Legal Cases Pertaining to Content Validation

- Guardians Association of New York City Police Department, Inc. v. Civil Service Commission of the City of New York (1980)
- Progressive Officers Club, Inc. v. Metropolitan Dade County (1990)
Conclusions on Content Validation (Stelly & Goldstein, 2007)

- The *Uniform Guidelines* erred in their distinction of construct and content; constructs are not entirely distinct from KSAOs.
- Content validation is suitable for selection procedures measuring constructs, if the constructs as not more abstract than necessary and have operational definitions (adequate methods of measuring them).
- Regardless of the validation strategy used, the critical question under Title VII is job-relatedness. Any exam used must be linked back to observable, on-the-job behaviors. If a trait or mental process is needed for performance, content validation can support the use of a selection procedure that tests for that trait.
Construct Validity

• Process of gathering evidence that the test is measuring what it claims to be measuring (e.g., general intelligence, conscientiousness, etc.)
  • Convergent/Discriminant Validity
  • Nomological Network
  • Multitrait-multimethod Matrix
Recommendations for Agencies Developing Exams In-House

- Always do a job analysis first and develop the exam plan based on the job analysis results
  - A general rule of thumb for updating job analyses is once every 5-7 years to be safe – however, in general job analyses should be updated as the requirements and tasks of the position change
  - Careful with technology and automation
- Do not rely **solely** on criterion-related validation – an inadequate or extremely flawed criterion validation study is not as legally defensible
  - If you’re feeling uneasy, consider hiring an I/O Psychology Practitioner
  - Establish content validity first, to be safe
- Do your research and consult your legal team before deciding strategy
Recommendations for Agencies Using Test Providers

- Be sure to review the test’s technical report for the job analysis information.
  - Even though the testing provider should have already conducted a job analysis, it is still important to gauge if the findings apply to your agency. If you already have a job analysis, you can do this by comparing your job analysis and results to that of the testing provider. If you do not already have an updated job analysis, you may try using the job analysis survey that should appear in the appendix of the test’s technical report and administering it to your subject matter experts.
• Look for testing providers that have gathered multiple sources of validity evidence
  • Test providers usually have better luck with larger sample sizes, but worse luck with criterion measures
  • Conducting your own local criterion-validation study would strengthen your legal support (and give you access to other criterion measures)
    • If possible, share the results of your study with the testing provider.
    • Check with your testing provider for services they offer regarding local criterion and validation
References and Resources

- Albemarle v. Moody, 422 U.S. 405 (1975)
- APA, AERA, and NCME’s Standards for Educational and Psychological Testing (2014)
- Equal Employment Opportunity Commission, Office of Personnel Management, Department of Justice, Department of Labor, & Department of Treasury (1979), Adopted of Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures. Federal Register, 44, 11996-12009.
• Guardians Association of the New York City Police Department, Inc. v. Civil Service Commission of the City of New York (1980), 630 F. 2d 79 (2nd Cir. 1980).

• Progressive Officers Club, Inc. v. Metropolitan Dade County (1990), 1990 WL 270786 (S. D. Fla.).

• Ricci v. DeStefano, 557 U.S. 557 (2009)


• Washington v. Davis, 426 U.S. 229 (1976)
Contact Yari

YRANDALL@IPMA-HR.ORG
WWW.LINKEDIN.COM/IN/YARI-RANDALL-581298127/
(703) 549 – 7100 EXT. 252