Selection Procedure Development and Validation

BCGi: Adverse Impact & Test Validation Book Series

Selection Procedure Development and Validation

• Selection Procedures that are properly validated should be both effective and defensible
• Content and Criterion-related Validity are two common and acceptable strategies for validating tests
• Standards for developing and validating tests include both Professional and Federal works
• Some strategies should be avoided in Title VII compliance or litigation settings

Why is this Topic Important to HR/EEO Professionals?

• Why do I need to know about this topic?
  – Validating your company’s tests will lead to a more qualified workforce and will help defend against lawsuits
  – FEA’s are regularly reviewing PPTs now
• What are the key essentials I need to know about this topic?
  – Employers cannot buy “valid” tests like they can “validated” products; if adverse impact exists, a test-job connection is required
  – Both professional and federal validation guidelines exist and are important to the process
• What are the consequences surrounding these issues?
  – The immediate consequence of using invalid tests is a less qualified and lower performing workforce
  – Need to properly calibrate tests to true job needs
  – The “start up” cost of a validation case is $30k to $80k
Presentation Overview

- Validation Defined
- Overview of the Mechanics of Content and Criterion-related Validity
- Benefits of the Validation Process
- Professional Standards for Validation
- Uniform Guidelines Requirements for Validation
- Blending the Professional and Government Validation Standards into Practice

Biddle Consulting Group, Inc.

- More than 30 years of experience in Equal Employment Opportunity consulting
- Has represented hundreds of employers in litigation-related settings
- Has performed job analyses and validation studies for hundreds of employers
- Has created valid selection testing that have been used by thousands of employers

This presentation is offered for information purposes only and should not be considered legal advice
What is Validation?

- Traditionally, validation is making sure a selection practice, procedure, or test (PPT) appropriately measures what it is designed to measure.
- In a legal realm, a selection procedure is valid if it can be proven by an employer that it is “…job related and consistent with business necessity.”

What to Validate

- Section 2 B of the federal Uniform Guidelines on Employee Selection Procedures indicates the Guidelines apply to tests and other selection procedures which are used as a basis for any employment decision. Employment decisions include but are not limited to hiring, promotion, demotion, membership (for example, in a labor organization), referral, retention, and licensing and certification, to the extent that licensing and certification may be covered by Federal equal employment opportunity law. Other selection decisions, such as selection for training or transfer, may also be considered employment decisions if they lead to any of the decisions listed above.

Test Validation

- You do not validate a “practice, procedure, or test.”
- You validate the inferences made from the results of those devices (e.g., test scores).
- Thus, you must specify how you intend to use a selection device for validity to be determined.
  - Validation looks at the use of a specific selection device for a specific purpose.
Why Validate?

- **It makes good business sense**
  - Validation often results in selection devices that are better at helping to choose the best people for the job
  - Can help to reduce the amount of time spent training newly-hired employees
  - Often helps to reduce turnover
  - Sends the message that you care about doing the “right thing”
  - Helps to minimize potentially-negative legal liability

- **It is required under some circumstances**
  - If there is adverse impact against a protected group of test takers (Executive Order 11246, As Amended)
  - For federal contractors if selection testing is identified as a problem during an audit

What is Adverse Impact?

- **A substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group**
  - **Title VII of the Civil Rights Act 1964**
  - **Also to the disadvantage of members who are those 40 or more years of age**
  - **Age Discrimination in Employment Act of 1967**

Adverse Impact

- **Determination of adverse impact relies on more than the “80% rule of thumb”**
  - 80% Rule of Thumb: When a selection rate for any race, sex, or ethnic group is less than eighty percent (four-fifths) of the selection rate for the group with the highest selection rate
  - The federal Uniform Guidelines refer to adverse impact in terms of differences in both statistical and practical terms
  - “Rather than using the 80 percent rule as a touchstone, we [the court] look more generally to whether the statistical disparity is ‘substantial’ or ‘significant’ in a given case” (Bozman v. Block, 1991, citing Contreas v. City of Los Angeles, 1981)
    - “Substantial or significant” refers to a statistically significant difference in passing rates between two groups (i.e., not likely occurring by chance)
  - **Practical significance tests evaluate the “practicality” or “stability” of the results**
    - Different courts have ruled differently on what is “practical”
    - BCG consultants can assist you in appropriately determining whether any differences detected are practically significant
The courts and auditing agencies are more likely to scrutinize validation studies that are conducted after adverse impact is determined.
How Can Testing Practices be Challenged? Title VII Disparate Impact Discrimination Flowchart

Is there Adverse Impact?

YES

Practice, Procedure, or Test (PPT)

NO

Is Practice, Procedure, or Test Valid?

Yes - Valid

NO - Defendant (Employer) Prevails

NO, not Valid - Complainant Prevails

How Can Testing Practices be Challenged? Title VII Disparate Impact Discrimination Flowchart

Is there an Alternate Employment Practice?

Complainant Prevails

Is there an Alternate Employment Practice?

Where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact.

Is there an Alternate Employment Practice?

Is there an Alternate Employment Practice?

NO, not Valid - Complainant Prevails

End of Story

Practice, Procedure, or Test (PPT)

Is PPT Valid?

Yes - Valid

NO - End of Story

Is PPT Valid?

Yes - Valid

NO - Complainant Prevails

Practice, Procedure, or Test (PPT)

Is there Adverse Impact?

YES

Is there an Alternate Employment Practice?

Complainant Prevails

Practice, Procedure, or Test (PPT)

Is there a PPT?

YES

Is PPT Valid?

YES - Valid

NO - Complainant Prevails

NO - End of Story

Practice, Procedure, or Test (PPT)

Is there a PPT?

NO - End of Story

Practice, Procedure, or Test (PPT)

Is there Adverse Impact?

YES

Is there an Alternate Employment Practice?

Complainant Prevails

Is there an Alternate Employment Practice?

NO - Defendant (Employer) Prevails

NO, not Valid - Complainant Prevails

Note: The defense of age-related adverse impact follows a different course than is shown here. Contact us if you would like more information about issues involving age-related adverse impact.
In order for there to be validity...

There must first be *reliability*

**Reliability:**

The degree to which test scores are consistent, dependable, or repeatable

The sun rising every day is a repeatable, reliable occurrence
Two Types of Commonly Considered Test Reliability

- **Internal consistency**
  - Looks at how well each test item relates independently to the rest of the items on the test and how they relate overall

- **Test – Retest reliability**
  - Test given more than once... people should score similarly each time they take the test
  - Acceptable method for those types of tests for which internal consistency is not appropriate, such as a typing test or work sample (e.g., finding a vein consistently to take blood)

There are other types of reliability that are less commonly used.

---

**Graphic Explanation of Reliability & Validity**

Not Reliable or Valid

- Scattered
Reliable... but not Valid

Consistent, but not on target

This reliably hits the target in the same place all of the time

But, it's the wrong part of the target!

Reliable and Valid

This reliably hits the target in the same place all of the time

Consistent, and on target

U.S. Department of Labor Reliability Guidelines

Table 1. General Guidelines for Interpreting Reliability Coefficients

<table>
<thead>
<tr>
<th>Reliability coefficient value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 and up</td>
<td>excellent</td>
</tr>
<tr>
<td>80 - 89</td>
<td>good</td>
</tr>
<tr>
<td>70 - 79</td>
<td>adequate</td>
</tr>
<tr>
<td>below .70</td>
<td>may have limited applicability</td>
</tr>
</tbody>
</table>

Validity

Federal Uniform Guidelines on Employee Selection Procedures Validation Types

- **Content Validity**
  - Section 14C and 15C
- **Criterion Validity**
  - Section 14B and 15B
- **Construct Validity**
  - Section 14D and 15D

www.uniformguidelines.com

A Brief Overview of Each Type of Validity

(in plain English!)

- **Content Validity**: A non-statistical connection between the important parts of the job and the test
  - Typically based on a Job Analysis and on ratings from Job Experts (Subject Matter Experts - SMEs)

- **Criterion Validity**: A mathematical study that shows the test predicts or is related to job performance:
  - A statistical study with results that are “statistically significant” (less than 5% by chance)

- **Construct Validity**: A connection between a test, a trait, and job performance:
  - “Empirical evidence” connecting the test to the trait and the trait to the job (requires both Content and Criterion Validity, plus more)
What Types of Validation are REALLY USED in Practice?

• Content Validity (80%)
• Criterion Validity (15%)
• Construct Validity (5%)

Why?
- Simplicity
- Defensibility

The courts have ruled that one type of validity is not necessarily better than another type of validity.

Content Validity

• Content-related Validity is demonstrated by data that the content of a selection procedure is representative of important aspects of performance on the job.
  – In other words, the content of the “test” is representative of the content of the job.

A Job Analysis is the basis for Content Validity

• A Job Analysis “analyzes” important work behaviors and identifies relative importance of those behaviors.
  – If behavior results in a work product, an analysis of the work product is conducted.
  – If work behaviors are not visible, we identify and analyze those aspects of the behaviors that can be observed and the observed work products.

• Those behaviors selected for measurement should be critical and/or important work behaviors constituting most of the job.

• It also identifies the knowledge, abilities, and other characteristics (KSAOs) needed to perform those behaviors (more about KSAOs to follow).
**What Ratings Does BCG Collect During a Content-related Job Analysis?**

- **For Job Duties** we collect ratings for:
  - Frequency* (Guidelines, 14.C[2])
  - Importance* (Guidelines, 14.C[2])
  - Differentiating* (Guidelines, 14.C[2])
  - Relative difficulty
  - Fundamental (Americans with Disabilities Act)
  - Assignable (Americans with Disabilities Act)

* = required under federal Guidelines
What Ratings are Collected during a Content-related Job Analysis?

- For knowledge, skills, abilities, and other characteristics (KSAOs) we collect ratings for:
  - Frequency* (Guidelines, 14.C[3])
  - Importance* (Guidelines, 14.C[3])
  - Performance Differentiating
  - Minimum v. Desirable Qualifications
  - Level needed upon entry
  - Level needed for success (knowledge only)

* = required under federal Guidelines

Targeting Key KSAOs for Test Building

- Only certain types of KSAOs can be tested when using content-related validity
  - The PPT measures and is a representative sample of a knowledge, skill, or ability and is
    - Used in and is a necessary prerequisite to performance of critical or important work behaviors (i.e., shown through a link to job duties)
    - Should closely approximate an observable work behavior, or its product should closely approximate an observable work product
    - If the PPT purports to sample a work behavior, the manner and setting of the PPT and its level of complexity should closely approximate the work situation
    - The closer the content and the context of the selection procedure are to work samples or work behaviors, the stronger is the basis for showing content validity
  - KSAOs being measured should be “operationally defined”

Content Validity

- A selection procedure based upon inferences about mental processes cannot be supported solely or primarily on the basis of content validity.
  - Content validity cannot support the measurement of traits (such as intelligence, aptitude, personality, commonsense, judgment, leadership, spatial ability)
- Not appropriate for those KSAOs that will be expected to be learned on the job or learned in a “brief orientation”
How do “Competencies” fit in?

Competencies

• There is no one, universally-accepted definition of a competency either legally or in the scientific literature
• Many say competencies are comprised of knowledge, skills, abilities, and other characteristics needed to successfully perform the job
  “Other Characteristics” in this case can range from personality traits, motives, values, self-concepts, job context/work environment, and more
• Important: Only knowledge, skills, and abilities that can be operationally defined in accordance with the federal Guidelines may be validated using a content-related approach

Content Validity Flow Chart
Criterion-related Validity

• Criterion-related Validity is demonstrated by empirical (observed) data showing that the selection procedure is predictive of, or significantly correlated with, important elements of work behavior.

For example: Criteria could include performance ratings, production rate, error rate, turnover/tenure, absenteeism, tardiness, disciplinary actions, etc.
We then compute the best fit of the relationship between test scores and the criteria.
**Strong, Positive Correlation**

Example: As test score goes up, productivity goes up.

**Weaker, Positive Correlation**

**Strong, positive correlation**

**Weaker, positive correlation**
Test Performance Criteria

Negative Correlation

Example: As test score goes up, number of absences goes down

Interpreting Correlation Coefficients

Correlations can range from +1.0 to -1.0

The closer to +1.00 or -1.0 the stronger the relationship between the variables

The stronger the relationship between two variables, the better the ability to predict one if given the other

U.S. Department of Labor

<table>
<thead>
<tr>
<th>Validity Coefficient Value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>above .35</td>
<td>very beneficial</td>
</tr>
<tr>
<td>.21 - .35</td>
<td>likely to be useful</td>
</tr>
<tr>
<td>.11 - .20</td>
<td>depends on circumstances</td>
</tr>
<tr>
<td>below .11</td>
<td>unlikely to be useful</td>
</tr>
</tbody>
</table>

* Many courts have ruled that .30 is the minimum acceptable validity if there is adverse impact against a protected group of test takers
* Uncorrected validity coefficients rarely exceed .40
Two Types of Criterion Studies

• Concurrent Study
  – PPT (e.g., test) is given to current employees
  – Job performance data is collected from same employees
  – Test scores correlated to job performance scores

• Predictive Study
  – PPT is given to job applicants, but scores are not typically used for selection purposes
  – Job performance data is collected from those job applicants who are hired
  – Test scores correlated to later job performance scores

Criterion Validation

• Can be a very powerful method for validating selection devices
• Usually used for tests that will be administered to extremely large groups of candidates and/or which will be used repeatedly
• Criterion-related studies often require 200 or more participants
• There is a chance that a significant relationship might not be found even if such a relationship actually exists

How helpful is a validity coefficient for understanding the relationship between test performance and job performance?

• It can be very useful when interpreted appropriately
  – However, meaningfulness can change based on whether the PPT is used pass/fail... rank ordered... other approaches. Each can cause the coefficient to mean something different
  – Industrial & Organisational Psychologists, such as those from BCG, can help you appropriately interpret these coefficients
• There are statistical tools that are used to “correct” validity coefficients
  – These can be easily misused or used incorrectly
  – Federal guidelines indicate that both uncorrected and corrected validity coefficients should be reported
Is Validation Forever?

• We recommend validation should be routinely reviewed every three to five years to determine whether it is still appropriate
• Re-validate whenever fundamental aspects of the job change or whenever the knowledge, skills, abilities, or personal characteristics needed to perform the job have changed

So... why don’t we only use tests that have no adverse impact?

• To insure successful job performance -- some abilities necessary for successful job performance are not distributed equally across groups
  – For example: A smaller proportion of women can perform physically demanding job tasks than men (Firefighters must be able to carry a heavy hose at the scene of a fire)
• To promote fairness to existing employees who are currently performing the job and to the job candidates who are applying for the job
  – Hiring candidates who cannot perform the job can negatively affect them and your current employees
Use selection tools that are fair to ALL applicants

This means...
- Discriminate only on a person’s ability to do the job
- Use only practices, procedures, and/or tests that can be linked to important aspects of job performance
  - Make sure you have the evidence to support claims of job-relatedness
  - You do not have to hire unqualified applicants

www.disparateimpact.com

Federal Standards and Adverse Impact: The Trigger

Requiring Validity for the Test and Use of the Test

- 2 X 2 Table Comparison
- Compares passing rates between 2 groups
- Probabilities < .05 = red flag
- Fisher Exact Test

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>Fail</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>Probability</td>
<td>0.200</td>
<td># SDs 1.280</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>Fail</td>
<td>50</td>
<td>65</td>
</tr>
<tr>
<td>Probability</td>
<td>0.045</td>
<td># SDs 2.000</td>
</tr>
</tbody>
</table>

In Summary...
Summary

- You must validate a selection device if there is adverse impact against a protected group of candidates.
- It is a good idea to validate even if there is no adverse impact.
  - The courts are more likely to scrutinize validity studies conducted after adverse impact is found.
- The courts have ruled that one type of validity is no better than another type of validity.

Neither Reliable or Valid

Reliable... but not Valid

Consistently missing the best performers
Reliable and Valid!

- Better Productivity
- Reduced Turnover & Absenteeism
- Higher Quality Output
- Employee Satisfaction Increased
- Less Time to Train
- Less Liability

Helpful Validation Resources

- Federal Uniform Guidelines on Employee Selection Procedures (free resource)
  - www.uniformguidelines.com
- U. S. Department of Labor’s Testing and Assessment: An Employer’s Guide to Good Practices (free publication)
  - www.onetcenter.org/dl_files/empTestAsse.pdf
- Adverse Impact and Test Validation: A practitioner’s guide to valid and defensible employment testing
  - Authored by Dan Biddle; available from BCG

Questions?