

Assessment Centers DO Have Validity to Measure Intended Constructs: Varieties of Validity Evidence

George C. Thornton III
Department of Psychology
Colorado State University

What is NOT Controversial:

- Prediction of managerial success and performance with the overall assessment rating
- Several meta-analyses show correlations ranging from .31 to .58 with variety of criteria

Criticisms of Assessment Centers

- Assessment centers do not have construct validity
- Assessment centers do not measure intended constructs

Basis for Criticisms is a Narrow Limited Set of Evidence

- Convergent and discriminant validity of
 - Within exercise dimension ratings
 - Post exercise dimension ratings

These data are based on the ratings of single dimensions rated in one exercise by one assessor

Problems with these conclusions:

- Even some of this evidence shows dimension effects are comparable to exercise effects
- This evidence is not based on assessment center ratings using the proper integration method --- the behavior reporting method
- **THERE IS MUCH OTHER EVIDENCE**

Varieties of Other Evidence

Ratings of dimensions show inter-rater agreement/reliability

Final dimension ratings correlate with ratings of comparable dimensions of job performance

Assessment center dimensions correlate with **OTHER METHODS** measuring comparable constructs

Final dimension ratings show more than halo

Traditional AC Method

- Multiple observers
- Observe overt behavior in
- Simulation exercises
- Classify behavior into performance dimensions
- [Assessors may or MAY NOT rate dimensions at this point]
- Integrate observations and evaluate dimensions and possibly overall performance

Convergent and Discriminant Validity of Within-Exercise Ratings

- Some evidence shows lack of discriminant validity
- Some analyses show exercise effects are larger than dimension effects
- Some factor analyses show exercise factors but not dimensions factors

- BUT

Other results of these types of studies

- Dimension effects are comparable to exercise effects
- Mean effect size for dimensions is .50 and mean effect size for exercises is .56

Analyses of Within Exercise Dimension Ratings are Limited

- The assessment center method stresses the value of aggregating across assessors, exercises, and dimensions
- Assessment center method places no value on the ratings of ONE assessor observing behavior on ONE dimension in ONE exercise

Original Behavior Reporting Method

- Does not call for rating dimensions by one assessor after a single exercise

Modern View of Validity

- “Validity is a unitary concept”

Validity is not separate things like:

Content validity

Criterion validity

Construct validity

Modern View of Validity

- “Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence... This includes:
 - Evidence of careful test construction
 - Adequate score reliability
 - Appropriate test administration and scoring
 - Accurate score scaling
 - Careful attention to fairness for all examinees.”

Varieties of Evidence According to *Standards*

- Test content
- Response processes
- Internal structure
- Relations with other variables
- Convergent and discriminant evidence
- Test-criterion relationships
- Validity generalization
- Consequences of testing

Test (Exercise) Content

- Simulations of tasks, content, and settings of target jobs
 - Individual exercises
 - Group exercises
 - Written exercises
 - Interaction exercises

Response Processes

- Assessors' observations and ratings of dimensions demonstrate inter-rater agreement and internal consistency
 - .75 to .89
- Ladd et al., 2002
- Schleicher et al., 2002

Final dimension ratings correlate with comparable dimensions on job

- Concurrent correlations with current managers ratings
- Predictive correlations with managers' rating several years after assessment

Dimension ratings correlate with OTHER METHODS measuring similar constructs

- Final dimension ratings correlate with
 - Tests measuring related constructs
 - Peer ratings of related constructs
 - Self ratings of related constructs

Assessment Center is the Method

- Contrast this with the view that exercises are the method

ACs show incremental validity over other assessment methods

- Incremental validity over
 - Tests of cognitive abilities
 - Questionnaires measuring personality traits

Final Dimension Ratings are not all halo

- Factor analyses of final dimension ratings reveal three to five clusters of factors

Positive Consequences of Assessment Centers

- Minimal differences across race, gender, and age groups
- Positive reactions to ACs
- Participation heightens readiness to learn
- Feedback leads to developmental planning
- Subsequent follow up actions lead to development

Conclusions

ASSESSMENT CENTERS DO HAVE
VALIDITY

ASSESSMENT CENTERS DO MEASURE
INTENDED CONSTRUCTS

INFERENCE ABOUT INTERPRETATIONS
OF ASSESSMENT CENTER RATINGS
SHOULD BE BASED ON A VARIETY OF
EVIDENCE