




Comments to Compliment and Complement Kevin Murphy on “Assessment Centers: Not As (in)Valid As You Think”


George C. Thornton III
Department of Psychology
Colorado State University
Fort Collins, Colorado 80523

- 
- **Compliment:** praise, admire, congratulate
 - **Complement:** something that makes whole or supplies mutual needs or lacks, either of two parts that mutually complete each other



Dr. Murphy

- Thank you attending Congress
- Provides a rare treat for participants
- We welcome skepticism and challenges
- Invite a dialogue
 - Selection of your comments
 - Alternative position by Thornton
 - Your reaction



“On the whole, studies are slightly discouraging”

- Thornton & Rupp, 2006: On the whole, diverse varieties of evidence support the validity of inferences from AC scores
- In your view, what is “discouraging” and what is “encouraging”?
- Is the glass “half empty” or “half full” ?

“strong exercise effects and relatively weak dimension factors”

- These conclusions come from analyses of post-exercise dimension ratings (PEDRs)
- In your view, what is the relative importance of evidence about PEDRs in comparison with evidence about:
 - Final dimension ratings?
 - Overall assessment ratings?



“rationale of AC...assess the same dimensions in several exercises”

- Yes, but it was never an assumption person will perform at the same LEVEL in separate exercises
- Would you allow for:
 - “Split rating” in final dimension rating, e.g., “He showed strong leadership in one-on-one but weak leadership in group situations”
 - Inconsistency, e.g., “She showed strong leadership on day one, but weak leadership on day two.”



“OARs are correlated in .25 - .35 range with performance ratings”

- Earlier meta-analyses show $r = .39$ with wide range of r ; better ACs show higher r
- Might it be that:
 - Earlier ACs were actually “better” ACs; better ACs show higher criterion validities ?
 - Research methods in the earlier studies were stronger, e.g., better criteria, wider range of abilities in assessees, etc. ?



“ACs are not worth the time and expense”

- Utility studies show benefits exceed expense; incremental validity is worth it
- What is an organization to do if the alternatives are not acceptable:
 - Larger adverse impact with traditional cognitive ability tests
 - The organization will not use personality questionnaires for fear of “faking”

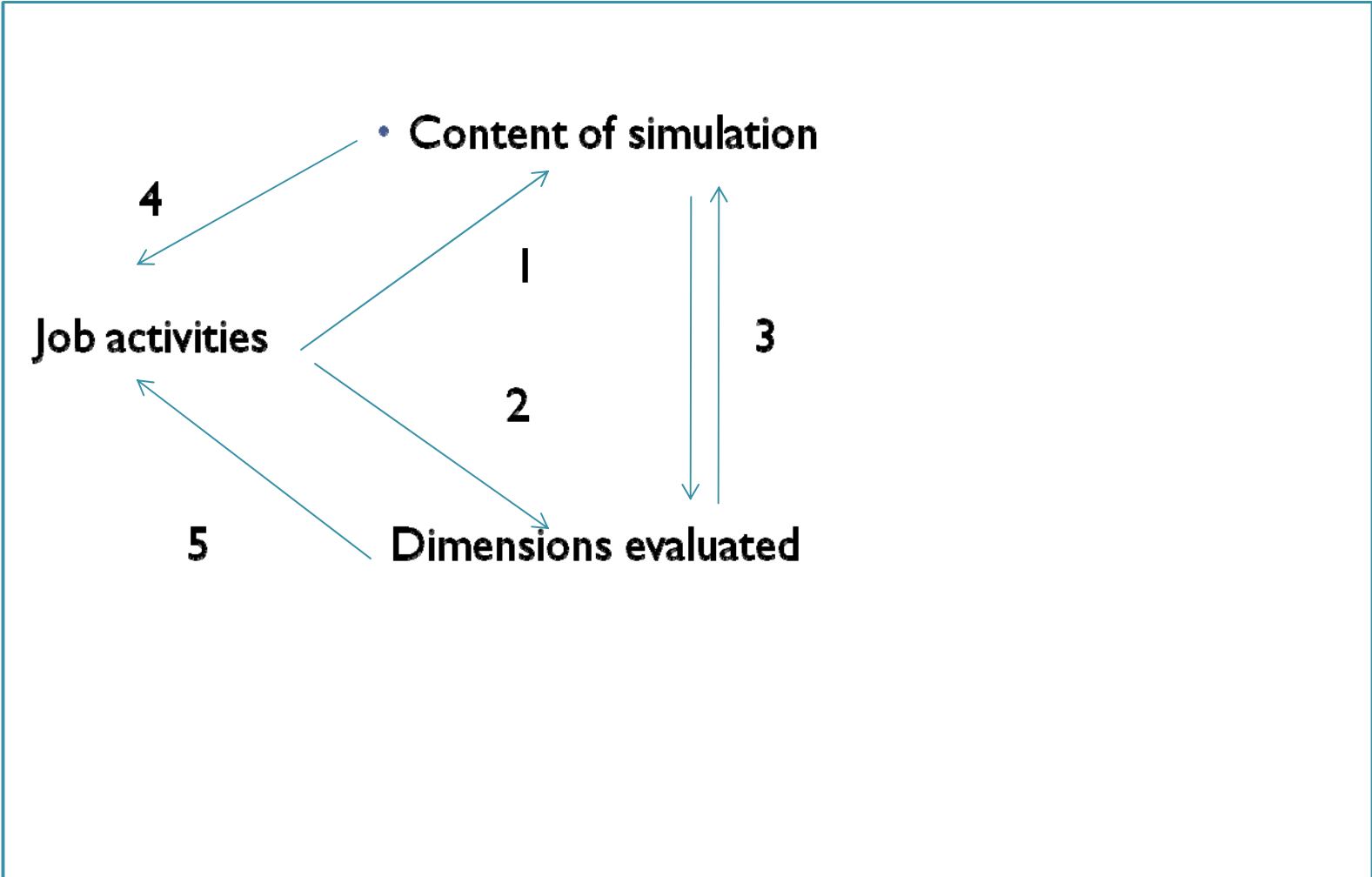



Content Validation Strategies: “Better than nothing but not much”

- I agree, IF content validation evidence is ‘The exercises look good’
- I do not agree, IF content validation includes evidence that:
 - Job requires Planning and Organizing, etc
 - Assessors can reliably observe and rate P&O in In-Basket, Group Discussion, etc

In your view, is any level of evidence of content representativeness alone adequate to demonstrate validity of AC?





- 
- Link 1: Content of exercise built to represent situations and problems in job
 - Link 2: Job analysis shows dimensions
 - Link 3: Behaviors shown in exercises fit with dimensions
 - Link 4: SMEs judge that content etc of simulations to be related to job
 - Link 5: SMEs judge dimensions to be important for job success



- In your view, is any level of evidence of content representativeness alone adequate to demonstrate validity of AC?



Summary

- What's an organization to do?