Construct validity controversy in AC research

- Common findings in AC research indicate that correlations between different dimensions within exercises typically exceed those of the same dimension across exercises (Sackett & Tuzinski, 2001).
- Often a MTMM approach is used to assess construct validity
  - Exercise vs. Dimension effects
MTMM approaches

- However, different MTMM models can yield different results
  - Lievens & Conway (2001) found roughly equal exercise and dimension effects using a CDCU model
  - Lance et al. (2004) found exercise effects four times as large as dimension effects using a one-dimension, correlated exercises model
  - Bowler & Woehr (2006) found stronger exercise than dimension effects, though not as large as Lance et al. using a CDCE model
Theoretical Models

CDCE Model
- Typical view of AC construct validity in which exercises and dimension affect post-exercise dimension ratings (PEDRs)
- Situational Specificity Hypothesis
  - Ratings represent valid perceptions of overall performance differences among AC participants
  - Performance is manifest consistently within an exercise, but may differ across exercises
  - No independent dimensions effects, only exercise effects
- Rating Error Hypotheses
  - General impression hypothesis
    - Raters form a general impression of respondents which then affects all dimension ratings within a given exercise (halo)
  - Salient dimension hypothesis
    - Assessors reach an exercise performance judgment based on a sole salient dimension, then this judgment affects other dimensions (spillover effect)

Recent work
- Recently, Lance, Woehr, & Meade (in press) simulated data under a variety of MTMM models:
  - CDCE – Typical view of construct valid AC
  - 1DCE – Indicative of situational specificity as well as SD and GI models
  - UDCE+g – General performance factor plus dimension specific and exercise factors
- Their results indicated that solution admissibility and model fit often can not be used to adequately determine which model reflects the true state of affairs
- Thus, a new way to assess construct validity is needed!
**Invariance constraints**

- Traditional measurement invariance research examines whether a construct is being measured equivalently across situations
  - Across cultures or other demographic groups
  - Over time
  - Across administration format
- Invariance constraints are also well-suited to examining whether AC dimensions are measured equivalently in different exercises

**Method**

- Sample was 298 managers enrolled in an Executive MBA program
- Four exercises
  - Two role plays
  - In-basket
  - Leaderless group discussion
- Four interpersonal performance dimensions
  - Oral Communication
  - Leadership
  - Confrontation
  - Sensitivity
- Three task performance dimensions
  - Analysis
  - Judgment
  - Decisiveness
Baseline Model

- Eight latent factors were formed that represented exercise by (task or interpersonal) performance dimension
- These are latent PEDRs for the four exercises and two performance dimensions
Research Questions:

1. Do AC dimension ratings within exercises contribute equally to latent PEDRs?
   - Assesses GI and SD hypotheses

2. Are AC dimension rating effects on observed PEDRs equal across exercises?
   - Does a given performance dimension have an equal effect across different exercise PEDRs
   - Assesses situational specificity hypothesis

Constrained model #1

- Factor loadings for all indicators that assessed the same latent PEDR were constrained to be equal
  - For example all factor loadings for the indicators of Task-Role Play 1 (Tsk_RP1) were constrained to be equal. Similar constraints were simultaneously placed on the other seven latent variables
Constrained model #1

- Tests whether different facets of task and interpersonal performance contribute equally to latent PEDR variable
- If constraints do not significantly worsen fit, support for the GI and SD hypotheses
- Note that dimensions being rated were similar (task or interpersonal)
Constrained model #1 Results

<table>
<thead>
<tr>
<th>Model</th>
<th>df (A)</th>
<th>$\chi^2$</th>
<th>$\Delta \chi^2$</th>
<th>ECVI</th>
<th>RMSEA</th>
<th>TLI</th>
<th>CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A: Baseline</td>
<td>271</td>
<td>808.43</td>
<td>--</td>
<td>1.96</td>
<td>.063</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>Model B: Constrained dimensions factor loadings within latent PEDR</td>
<td>293 (22)</td>
<td>1194.84</td>
<td>386.41*</td>
<td>2.65</td>
<td>.079</td>
<td>.89</td>
<td>.90</td>
</tr>
</tbody>
</table>

- Model fit significantly worse, indicating that the strength of the relationship between the latent PEDR and performance dimensions differs by dimension.
- Does not support GI or SD hypotheses

Constrained Model #2

- Factor loadings for all indicators that assessed the same indicator-level performance dimension were constrained to be equal across PEDRs.
Constrained Model #2

- As each latent PEDR is exercise specific, this is a test of whether the specific dimension are equally observed across exercises.
- Significantly worse fit provides some support for dimension effects.
- Significantly worse fit would mean that different performance dimensions are manifest differently in the exercises.
Constrained Model #2 Results

- Model fit significantly worse
  - Dimension effects are differentially manifest across exercises
- Follow-up tests constrained parameters for one specific indicator level dimension at a time to determine source of misfit.

<table>
<thead>
<tr>
<th></th>
<th>df (Δ)</th>
<th>$\chi^2$</th>
<th>$\Delta\chi^2$</th>
<th>ECVI</th>
<th>RMSEA</th>
<th>TLI</th>
<th>CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A: Baseline</td>
<td>271</td>
<td>808.43</td>
<td>--</td>
<td>1.96</td>
<td>.063</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>Model C: Constrained dimensions across exercise PEDR</td>
<td>288 (17)</td>
<td>887.06</td>
<td>78.63*</td>
<td>2.05</td>
<td>.065</td>
<td>.92</td>
<td>.93</td>
</tr>
</tbody>
</table>

Follow-up test results

<table>
<thead>
<tr>
<th></th>
<th>df (Δ)</th>
<th>$\chi^2$</th>
<th>$\Delta\chi^2$</th>
<th>ECVI</th>
<th>RMSEA</th>
<th>NNFI</th>
<th>CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A: Baseline</td>
<td>271</td>
<td>808.43</td>
<td>--</td>
<td>1.96</td>
<td>.063</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>Oral Communication</td>
<td>273 (2)</td>
<td>808.80</td>
<td>.37</td>
<td>1.95</td>
<td>.063</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>Analysis</td>
<td>274 (3)</td>
<td>814.63</td>
<td>6.20</td>
<td>1.96</td>
<td>.063</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>Judgment</td>
<td>274 (3)</td>
<td>815.06</td>
<td>6.63</td>
<td>1.96</td>
<td>.063</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>Decisiveness</td>
<td>273 (2)</td>
<td>826.15</td>
<td>17.72*</td>
<td>1.99</td>
<td>.064</td>
<td>.92</td>
<td>.93</td>
</tr>
<tr>
<td>Leadership</td>
<td>273 (2)</td>
<td>821.31</td>
<td>12.88*</td>
<td>1.98</td>
<td>.063</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>Confrontation</td>
<td>273 (2)</td>
<td>808.80</td>
<td>.37</td>
<td>1.95</td>
<td>.063</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>273 (2)</td>
<td>823.86</td>
<td>15.43*</td>
<td>1.98</td>
<td>.064</td>
<td>.92</td>
<td>.94</td>
</tr>
</tbody>
</table>
Follow-up test interpretation

- Significantly worse fit for the decisiveness, leadership, and sensitivity dimension constraints
- No loss of fit for constraints placed on the oral communication, analysis, judgment, and confrontation dimensions

Summary of Results

1. No support for salient dimension and general impression hypotheses.
   - Halo effects do not appear to be solely responsible for lack of dimension effects.
2. There are dimension effects, however weak they may be.
3. Performance dimension effects manifest themselves with varying degrees of strength in different exercises. We call this “differential manifestation”.
**Discussion**

- Invariance constraints provide an alternate approach to examining the construct validity of AC dimensions
- Provides an indirect test of previous explanations for the impact of exercises and dimensions on AC ratings

---

**Implications for CDCE Model**

- Same dimension rated in different exercises were equivalent for half of the dimensions
  - For these dimensions, the observed PEDR is invariant to the exercise in which the performance is observed
  - For the others, the exercise plays a role on the behaviors displayed
- Provides partial support for dimension-based interpretation of ACs
Implications for Situational Specificity Hypothesis

- Different dimensions rated in the same exercise were not equivalent
  - Thus, what is being rated in the exercise is not necessarily a unitary phenomenon
  - Distinct construct being rated (e.g., performance is not consistent within a given exercise)
- Consistent with “differential manifestation”

Conclusion/Implications

- Consensus discussions
  - In some cases, arriving at consensus for the same dimension assessed using different exercises is akin to talking “apples and oranges”
  - Consensus discussions should only focus on those dimensions that were rated equivalently across exercises
- Focus of Feedback
  - As opposed to dimension based feedback, feedback should be PEDR specific when possible