

# **Measuring Change in a Developmental Assessment Center: Evidence for Construct Validity**

**Deborah E. Rupp, PhD**

Department of Psychology,  
Institute of Labor and Industrial Relations  
University of Illinois at Urbana-Champaign

**George C. Thornton III, PhD**

Department of Psychology  
Colorado State University

**Alyssa Mitchell Gibbons, MS**

Department of Psychology  
University of Illinois at Urbana-Champaign

# Overview

- I. Getting on the same page:  
Developmental assessment centers  
(DACs) as interventions
- II. The Rupp & Thornton DAC Model.
- III. The Model in Practice: The Illinois  
Managerial Development Program
- IV. Validity evidence from Year 1 of  
operation
- V. Challenges
- VI. Research needs

# I. Getting on the same page: DACs as interventions

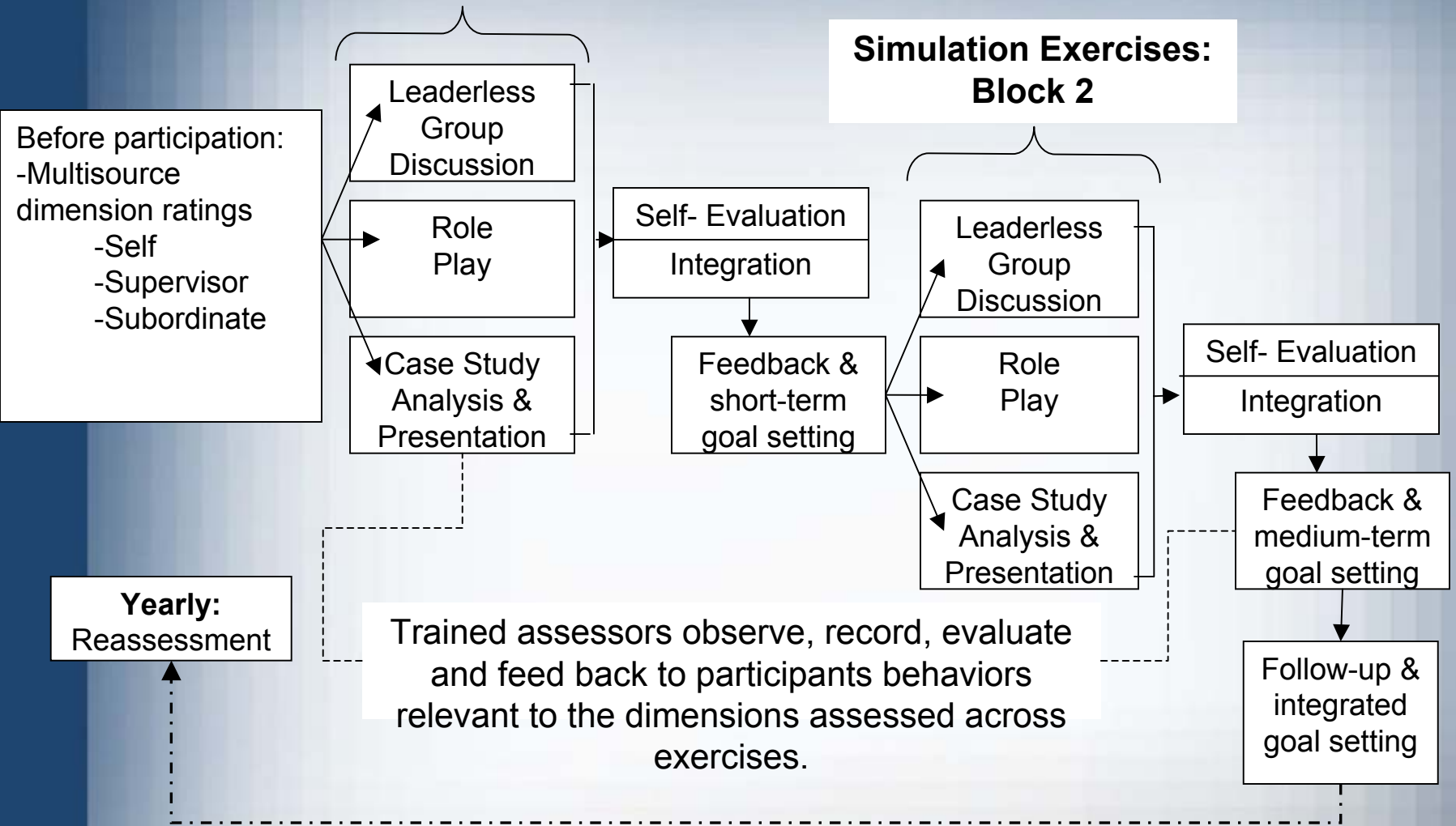
- Assessment centers (ACs) are increasingly used by organizations for *developmental purposes* (Kudisch et al., 2001; Spsychalski et al., 1997; Thornton & Byham, 1982).
- ACs have long been used for diagnosis of developmental needs (Lievens & Klimoski, 2001; Thornton, 1992), but...
- DACs may also catalyze development and can serve as an integral part of the development process (Carrick & Williams, 1999; Jones & Whitmore, 1995; Rupp & Thornton, 2003; Thornton & Rogers, 2001; Thornton & Rupp, in press).
- Although research is needed!

# II. The Rupp & Thornton DAC Model

(inspired by Boehm, 1985; Carrick & Williams, 1999; Engelbrecht & Fischer, 1995; Goodge, 1997; Griffiths & Goodge, 1994; Jones & Whitmore, 1995; Lee, 2000, 2003; Lee & Beard, 1994; Povah, 1986)

## Simulation Exercises: Block 1

## Simulation Exercises: Block 2



### III. The Model in Practice: The Illinois Managerial Development Program

- Longitudinal research project examining the validity of DACs.
- Open to mid-level managers from any organization.
- Purely developmental: feedback is given **ONLY** to the participant.

# MDP Dimensions

- Information Seeking
- Planning & Organizing
- Problem Solving
- Oral Communication
- Leadership
- Conflict Management
- Fairness
- Cultural Adaptability

# Feedback

- DAC Morning Feedback
  - Strengths and developmental needs on dimensions in morning exercises
  - Goal setting to improve in afternoon
- DAC Afternoon Feedback
  - Strengths and developmental needs in afternoon exercises
  - Observed improvement/goal attainment from morning to afternoon
  - Transfer of training—goals to continue developing when back on the job
- Post DAC
  - Multisource feedback on dimensions (level of proficiency on dimensions from assessors, self, supervisor, and subordinates)
  - Feedback on personality and vocational interests
  - Developmental tips for improving in areas of needs
  - Suggested readings
  - Written report plus follow-up session with facilitators

# Participants – Year 1

- 100 managers from Midwestern organizations:
  - banking, electronics manufacturing, state regulatory, research consultancy, county government, utility.
- 69% described themselves as middle managers.
- 50% female, 91% Caucasian
- Average experience:
  - 4.47 years in present position
  - 7.61 years with organization
  - 20.28 years in workforce

# MDP Innovations

- Comprehensive/innovative training program
- Experimenting with computerized assessments
- Multi-item ratings of dimensions (BARS).
- Computerized mobile assessor work stations.
- Streamlined computer-aided integration process.
- On-line multi-source ratings system

# Computerized Assessor Interface

WL LGD PS Frame : Form

**WL Leaderless Group Discussion - Problem Solving**

Jane Doe  
999999  
7/26/2004

**Problem Understanding**

Records of Evidence: + Clearly identifies important information in the text, as well as missing information.

1 2 3 4 5 6 7

Makes multiple factual errors about policies, employees, or exercise instructions

Draws group into discussion of irrelevant issues, such as how to cut costs in other areas

Differentiates between relevant and irrelevant information in profiles

Identifies "make or break" factors and weights other info appropriately

**Thinking Solutions Through**

Records of Evidence: + provides support for reasoning (i.e., "we can't fire John because he wants someone younger...")

1 2 3 4 5 6 7

Jumps to conclusions and refuses to be swayed by other information

Provides good reasons for own suggestions

Differentiates between relevant and irrelevant information in profiles

Identifies potential future (pos and neg) consequences of losing each employee

**Decisiveness**

Records of Evidence: + Commits to her decision and opinions.  
- Needs to reconfirm and enforce it toward the end of LGD.

1 2 3 4 5 6 7

Delays group decision by refusing to decide on one or more employees

Avoids stating own opinion, even when asked by others

Commits to opinions – "I think we should do X"

Encourages a decision when (and only when) all relevant information has been discussed

**Overall Problem Solving Rating**

1 2 3 4 5 6 7

Next Dimension

# Computer-Aided Integration Process



MDP Afternoon Report for: Jane Doe

10/4/2004

---

## Information Seeking

CP: Mary Brown

IS: Lisa Jones

---

### Use of Multiple Sources

CP Rating:

6

IS Rating:

5

LGD +Clearly express the desire to collect info from different perspectives/sources

IS: + ask good open ended questions of both parties  
+ asked more specifics  
+ got to the underlying issues

---

### Situational Relevance

CP Rating:

7

IS Rating:

6

LGD +Makes a clear chart listing pro's and con's of each option

IS: + got to underlying issues  
+ asked good [fu] questions to gain understanding

---

### Creation of Usable Patterns

CP Rating:

6

IS Rating:

6

LGD +Did mention culture issue at the beginning  
-But not in the {corporation}

IS: + validated information on both sides and brought it together when they came together  
+ set up a framework and solution

---

## IV. Validity Evidence: Year 1 of Operation.

- Validity requires many types of evidence:
  - A. Content
  - B. Internal structure
  - C. Convergent & discriminant relationships
  - D. Test-criterion relationships
  - E. Consequences of testing

*AERA/APA, 1999;  
Thornton, yesterday;  
Thornton & Rupp, in press*

# A. Content Validity

- All ACs require dimensions that are job-relevant (*International Taskforce, 2000*).
- DACs also require dimensions that are developable.
  - Objectively developable.
  - Perceived as developable by participants.
    - Beliefs about developability affect development behavior (*Dunning, 1995; Dweck & Leggett, 1988; Zuckerman, Gagne, and Nafshi, 2001*).
- Some dimensions may be more developable than others (*Rupp et al., 2003*).

# Potential Continuum of Developability

	<b>Nearly impossible to develop</b>	<b>Very difficult to develop</b>	<b>Difficult to develop</b>	<b>Reasonable possibility to develop</b>	<b>Somewhat easy to develop</b>
<b>Dimension</b>	Motivation	Adaptability Conscientiousness	Interpersonal Skills	Listening Problem solving techniques	Nonverbal communication
<b>Development Strategies</b>		Long-term practice with coaching Extensive counseling	Counseling Education programs Long training programs Courses Mentoring Skill Practice	Feedback alone Lectures/seminars Readings On the job experience with coaching	Participation in a simulation alone Self insight

*Rupp et al. (2003)*

# Content Validity Evidence

- Dimensions chosen based on a synthesis of general taxonomies of managerial performance dimensions (e.g., *Borman & Brush, 1993; Gaugler, Bentson, & Pohley, 1990; Tett, Guterman, Bleier, & Murphy, 2000*).
- Some empirical evidence exists that most of these dimensions can be developed by some method (e.g., *Fedorko, 1992; Kho, 2001; Roland, 1998; Smith & Smith, 1995*).
- Tested perceived developability via surveys in two cultures: US and South Korea (*Rupp et al., 2003*).

# Participants' Perceptions of Content

- From anonymous post-program evaluations (N=44):
  - Agreement = rating > 4 on a 7-point scale.
  - 71% agreed that the **exercises** were **appropriate** for their jobs.
  - 96% agreed that the **dimensions** were **relevant** to their jobs.
  - 89% thought the exercises were at an **appropriate level of difficulty**
  - 93% found the role players believable.

## B. Internal Structure

- The validity of assessment center ratings has often been questioned (*e.g., Lance et al., 2004; Sackett & Tuzinski, 2001*).
  - Dimensions are highly correlated within exercises.
  - Possible explanations: Halo, situational specificity.
- Challenge for all ACs, but especially DACs
  - need valid ratings in order to give valid and meaningful feedback. (*Thornton & Rupp, in press*)

# New Approach to Internal Structure

- Each dimension divided into three subdimensions.
  - Simplifies complex dimensions for assessors.
  - Provides multi-item measures of each dimension **within** each exercise.
- Within-exercise analyses allow us to ask whether assessors differentiate between dimensions.

# Within-Exercise Factor Analyses

- Compared single-factor model (halo) with five-factor model (five dimensions assessed).
- For all six exercises, the five-factor model showed a significant improvement in fit over the single-factor model.
- Suggests assessors are distinguishing between dimensions within the exercise!

# Example CFA Results:

## Group Discussion Exercise 1

	<b>Single Factor</b>	<b>Five Factor</b>	<b>Change</b>
$\chi^2$	158.94	120.31	$\Delta \chi^2 = 38.63$
df	90	80	$\Delta df = 10$
$\chi^2 / df$	1.77	1.50	$p < .01$
RMSEA	.10	.08	
NFI	.83	.87	
TLI	.91	.94	

Conservative example -  $\Delta \chi^2$  for other exercises was as high or higher.

# Dimension Correlations

## Morning Block

	IS	PO	PS	OC	LS	CM	F
Information Seeking	1.0						
Planning/Organizing	.69	1.0					
Problem Solving	.54	.58	1.0				
Oral Communication	.46	.53	.75	1.0			
Leadership	.57	.56	.65	.64	1.0		
Conflict Management	.52	.56	.58	.56	.78	1.0	
Fairness	.63	.66	.72	.59	.63	.67	1.0

# C. Convergent and Discriminant Relationships

- Dimensions that are logically similar should be more highly correlated than dimensions that are different. (*Shore et al., 1990, 1992*)
- Shore et al. (1990, 1992) classified dimensions into “performance” and “interpersonal.”
- Only partially supported by MDP data:

	<u>Block 1</u>	<u>Block 2</u>
– All dimensions:	$r = .61$	$r = .61$
– Performance:	$r = .59$	$r = .63$
– Interpersonal	$r = .70$	$r = .67$
- Also external approach—still sorting out those data

## D. Test-Criterion Relationships

- “Criterion” for DACs is improvement on dimensions over time.
- Longitudinal measures are needed.
  - Pre- and post-participation assessment
  - Reassessment
  - Follow-up
- Reliable external measures are needed.
  - Multisource ratings (*Engelbrecht & Fisher, 1995*)

# Conceptualizing Change

- Many opportunities for change to occur:
  - Within-program change
    - Objective (perceived by assessors)
    - Subjective (self-perceptions)
    - Active learning (involvement, trying new behaviors)
  - Post-program change
    - Multiple perspectives: self, supervisor, subordinates
    - Compared to pre-program performance
    - Compared to non-participants' development
    - Gamma change (*Rogers, in progress*)

# Objective Within-Program Change

- Comparison of assessor ratings for morning and afternoon blocks:
  - Repeated-measures MANOVA
  - Time (pre vs. post) = repeated-measures independent variable; Dimensions = measures
  - $F_{6, 91} = 4.97, p < .01$
  - Significant univariate effect for Oral Communication.

# Perceived Within-Program Change

- Comparison of participants' self-critiques within the program – morning and afternoon self-ratings.
  - MANOVA:  $F_{6, 89} = 2.52, p = .03$
  - Significant univariate effect for Oral Communication.
- Item: “Did your performance change from the morning to the afternoon? In what way?”
  - Response scale: 1 = “better in the morning” – 4 “about the same” – 7 “better in the afternoon”
  - Significantly greater than 4 for Information Seeking, Leadership, and Oral Communication.

# Active Learning

- Item: “In the afternoon sessions, did you try anything new or unusual for you with respect to this dimension?”
- For each dimension, at least 52% of respondents said yes.
- All but 5 participants said yes for at least one dimension.
- Average # of dimensions on which new things were tried: 3.26

# Post-Program Improvement

- Looked at self, supervisor, and subordinate ratings separately (different Ns).
- Subordinate ratings were averaged across all subordinates providing ratings for the same participant.
- Each dimension rating based on 3 items– equivalent to the subdimensions used by assessors.

# Post-Program Improvement – Self Ratings

Overall MANOVA:  $F_{8, 26} = 2.94, p = .02$

	Pre	Post
Information Seeking	5.48	5.59
Planning/Organizing	4.96	5.15
Problem Solving	5.37	5.56
<b>Oral Communication</b>	<b>4.89*</b>	<b>5.34*</b>
Leadership	4.96	5.23
<b>Conflict Management</b>	<b>4.66*</b>	<b>5.17*</b>
<b>Cultural Adaptability</b>	<b>5.46*</b>	<b>5.83*</b>
<b>Information Seeking</b>	<b>5.30*</b>	<b>5.86*</b>

# Post-Program Improvement – Multisource Ratings

- No significant change observed by supervisors or subordinates:
  - MANOVA:  $F_{8, 16} = 0.54, p = .81$
  - MANOVA:  $F_{8, 40} = 0.53, p = .83$
- Not panicking yet:
  - Much less data to date (low power)
  - No incentive, no control
  - These source have not been trained on dimensions & their scaling!
  - Year 1 participants not given 360 feedback prior to post participation 360 data collection
- Ultimate test: change in dimension proficiency when reassessed via the DAC one or more years later (we have just begun our first phase of reassessment)

## E. Consequences of Testing

- ACs are frequently perceived as fair and unbiased assessments.
  - Hopefully also the case for DACs – needs to be established.
- DACs need to be perceived not only as fair but as helpful.
  - Participant satisfaction is important.
  - Participants' intent to use feedback afterward is more so.

# Social Consequences

- ACs have generally been found to be fair to members of minority groups (*e.g.*, Goldstein, Yusko, & Nicolopoulos, 2001; Huck & Bray, 1976; Shore et al., 1997 )
- Questions have been raised about whether ACs may show age or gender bias (*Bobrow & Leonards, 1997; Clapham & Fulford, 1997; Dulewicz & Fletcher, 1982; Shore, 1992*).
- We found no evidence of age or gender bias in the MDP.
  - No effects of age or gender on dimension ratings.
  - Race could not be tested as sample was 91% Caucasian.

# Participant Reactions

- From anonymous post-program evaluations:
  - 100% agreed that the feedback seemed **accurate**.
  - 96% thought the **feedback** was **helpful**.
  - 89% believed that they were **more aware** of their development needs as a result of participating.
  - 95% found the **program helpful**.
  - 98% would **recommend it to others**.

# Participant Reactions

- 95% stated that they had **tried at least one new behavior** during the program.
- 98% said they had **identified at least one dimension they needed to develop**.
- 100% said they had **set at least one short-term development goal**.
- 77% **planned to talk to a supervisor** soon about their development plans.

# Summary of Validity Evidence

- All in all, this provides some preliminary evidence that the DAC process model proposed here:
  - Allows for the assessment of relevant dimensions, perceived to be developable (content)
  - assessors can discriminate between dimensions (internal structure)
  - Assesseees show & perceive improvement within and following DAC participation on dimensions proposed to be more developable (criterion relations)
  - Positive consequences result

# V. Challenges

- Still only a preliminary sample – are continuing to partner with new organizations
- Collecting multisource ratings from several different organizations is difficult.
  - Implementing online data collection – easier for users and allows better data management.
- Multisource ratings can be problematic
  - Little instruction and no training given to raters.
  - Requires raters to decide what “proficient” means.
  - Rating is voluntary – may not be representative.
  - Multiple motives to provide ratings.

## VI. Research Needs

- More detailed longitudinal follow-up
  - When does development occur?
- Investigation of participants' development behavior post-program.
- More direct investigation of the impact of developability beliefs.
- More measurement of external variables to test convergent/discriminant relationships.
- Further investigation of the structure of assessor ratings.

# END...Thank you!

For more information or to  
provide your feedback:  
[derupp@uiuc.edu](mailto:derupp@uiuc.edu)

# Summary of Validity Evidence

- **Content Validity:**
  - The dimensions used in the MDP are perceived as job-relevant and (at a general level) developable.
  - Participants viewed the exercises as job-relevant and at an appropriate level of difficulty.
- **Internal Structure:**
  - MDP assessors appear to differentiate between dimensions within exercises.
  - A hierarchical model (general factor + dimensions) might fit even better, but the present sample is insufficient to test it.

# Summary of Validity Evidence

- Convergent/Discriminant Relationships:
  - Classifying dimensions based on the Shore et al. (1990, 1992) model was only partially successful.
  - Interpersonal dimensions (leadership, conflict management, fairness, cultural adaptability) were more highly correlated with each other than with performance dimensions.
  - Performance dimensions (planning & organizing, information seeking, oral communication, problem solving) were no more correlated with one another than with any other dimension.

# Summary of Validity Evidence

- Change:
  - Assessors observed improvement in participants' Oral Communication performance within the program.
  - Participants perceived similar improvement on Oral Communication in their own self ratings.
  - Participants also felt their performance improved for Information Seeking and Leadership.
  - Most participants reported trying a new behavior or approach (engaging in active learning).
  - Trying something new had a positive relationship with performance for several dimensions.
  - After the program, participants perceived improvement in their own performance on four of eight dimensions.

# Summary of Validity Evidence

- Consequences:
  - The MDP shows no evidence of bias with respect to age or gender.
  - Participant reactions to the program were very positive.
  - Participants indicated positive behavioral intentions to pursue development after the program.