

## Using Item Response Theory to Assess Measurement Equivalence of 360° Performance Ratings Across Organizational Levels

S. Bartholomew Craig  
North Carolina State University

Robert B. Kaiser  
Kaplan DeVries Inc.

In A. Meade (Chair) *Applications of IRT to Measurement in Organizations*.  
Symposium presented at the annual conference of the Society for  
Industrial-Organizational Psychology, Orlando, Florida, April 11, 2003.

## How to Get the Paper

- ▶ Download from [www.sbcraig.com](http://www.sbcraig.com)
- ▶ Request a copy via e-mail from [bart\\_craig@ncsu.edu](mailto:bart_craig@ncsu.edu)
- ▶ Leave your contact information after the presentations

## Background

- ▶ 360° performance ratings have become the norm in organizations
  - § Primary value lies in opportunity to compare ratings from different sources
  - § Comparison implicitly assumes measurement equivalence

## Background

- ▶ Research has only recently begun testing the assumption of measurement equivalence
  - § Race
  - § Language translations
  - § Mode of administration (paper vs. electronic)
  - § Geographic distance
  - § Time
  - § Rater – ratee relationship (rating source)

## Crossing Organizational Levels

- ▶ Previous research on equivalence of rating sources does not address organizational levels
- ▶ Theoretical bases for possible inequivalence across levels
  - § Differing skill requirements (Katz, 1955; Mann, 1965)
  - § Application of structure (Katz & Kahn, 1978)
  - § Interactive Complexity Theory (Streufert & Swezey, 1986)
  - § Stratified Systems Theory (Jacobs & Jaques, 1987)
- ▶ Need for new research that disentangles rating source from organizational level

## Method

- ▶ Archival database of *Benchmarks*<sup>®</sup> ratings
  - §  $N = 85,342$  raters / 8434 ratees
  - § Only peers with unambiguous level used ( $N = 3783$  raters / 974 ratees)
  - § Response scale collapsed to four points
- ▶ EFA used to empirically derive factors
- ▶ CFA used to test between-factors equivalence
- ▶ DFIT used to test within-factor equivalence

## Results—EFA

- ▶ EFA suggested 3 factors (83% of variance)
  - § *Decisive Action Orientation*
    - ▶ "is action oriented"
  - § *Relationship Orientation*
    - ▶ "has a warm personality that puts people at ease"
  - § *Working Through Others*
    - ▶ "focuses more on managing other people to accomplish a task than on personally finishing everything the workgroup does"
- ▶ Highest loading 4 items from each factor used in subsequent analyses

## Results—CFA

	CFI	NNI	RMSEA	$\chi^2 / df$
All Peers ( <i>N</i> =26,536)	.95	.93	.075	149
Lowest Level ( <i>N</i> =617)	.94	.93	.075	4.46
Middle Level ( <i>N</i> =2211)	.95	.94	.072	12.4
Highest Level ( <i>N</i> =348)	.94	.92	.079	3.15

## Results—DFIT

- ▶ All possible pairwise comparisons among levels were tested (3)
- ▶ 36 item-level tests conducted (NCDIF)
  - § None was significant
- ▶ 9 scale-level tests conducted (DTF)
  - § None was significant

## Interpretation

- ▶ Conclusions
  - § These 3 scales function equivalently at each of these three organizational levels
  - § These 3 scales can be meaningfully compared
- ▶ Limitations / Future Research
  - § Constructs not explicitly chosen to reflect dimensions thought to vary across levels
  - § Small number of constructs
  - § Small number of items per construct