Beyond Measurement Artifacts: Integrating Measurement Equivalence with Theory Development in Cross-Cultural Research

Gordon W. Cheung, Ph.D

Professor of Organizational Behaviour
Department of Management and International Business
The University of Auckland Business School
Past Division Chair, Research Methods Division
Academy of Management

1

What is Measurement Equivalence/Invariance (ME/I)

ME/I is a general term that can be applied to the comparison of the various components of measurement models, and can sometimes be extended to structural models and mean structures

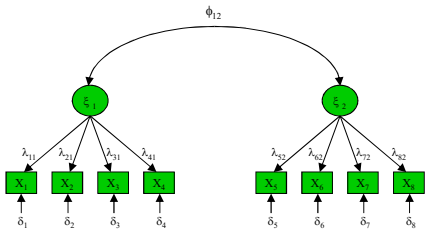
2

Basic Types of ME/I

- Configural Equivalence
- Metric Equivalence
- Scalar Equivalence
- Uniqueness Equivalence
- Construct Variance Equivalence
- Construct Relations (Covariance and Path Coefficients) Equivalence
- Latent Mean Equivalence

3

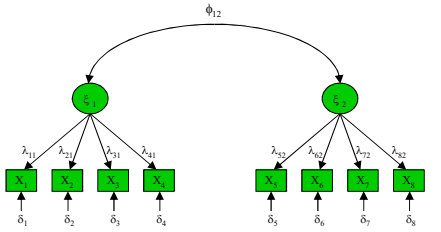
Measurement Model



4

Configural Equivalence

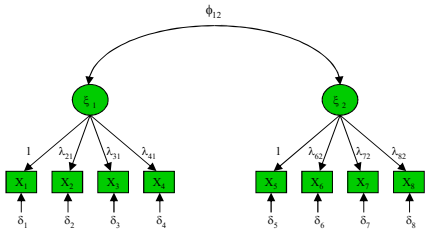
- All groups associate the same subsets of items with the same constructs (the cognitive domains are the same)



5

Metric Equivalence

- Overall, the strength of the relationships between items and their underlying constructs are the same across groups. (The constructs are manifested in the same way)



6

Scalar Equivalence

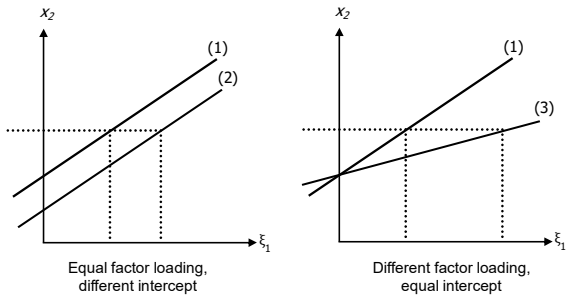
- Intercepts are the same across-groups. The cross-group differences indicated by the items are the same across items. Alternatively: all items indicate the same cross-group differences.

$$x_i = \tau_i + \lambda_{ij}\xi_j + \delta_i$$

7

Metric and Scalar Invariance

$$x = \tau_x + \Lambda_x\xi + \varepsilon$$



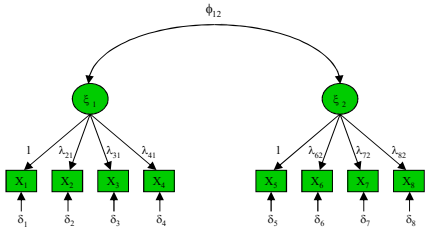
Equal factor loading, different intercept

Different factor loading, equal intercept

8

Uniqueness Equivalence

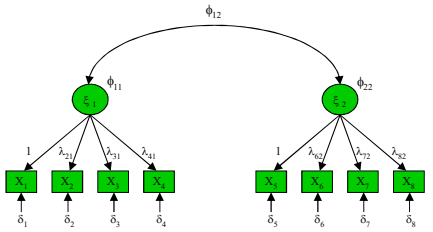
- Items demonstrate the same size of measurement error across groups. Alternatively: Items have the same quality as measures of the underlying construct across groups.



9

Construct Variance Equivalence

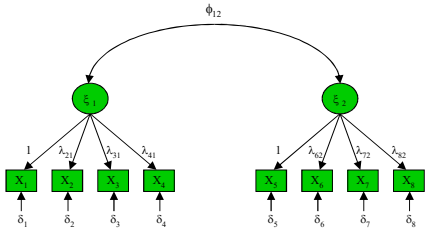
- The range of responses given to each item is the same across groups. Alternatively: the variability / range of diversity with respect to the constructs are the same across groups.



10

Construct Covariance Equivalence

- The relationships among constructs (e.g., covariance and regression coefficients) are the same across groups.



11

Latent Mean Equivalence

- The mean level of each construct is the same across groups.

$$\boldsymbol{E}(\boldsymbol{x}_i) = \boldsymbol{\tau}_i + \boldsymbol{\lambda}_{ij}\boldsymbol{\kappa}_j$$

$$\boldsymbol{\kappa}_j^{(1)} = \boldsymbol{\kappa}_j^{(2)}$$

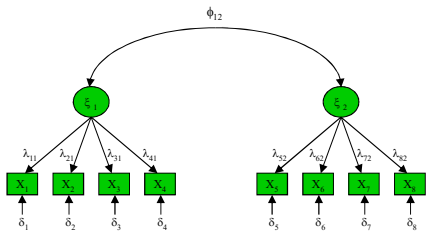
12

Applications of Multi-Group Analysis

- Independent Group Model: Cross-cultural comparisons of job satisfaction
- Non-independent Group Model: Disagreement in multi-source performance appraisal
- Longitudinal Model: Revisiting the Alpha, Beta, Gamma Change Typology

13

Cross-Cultural Comparison of Job Satisfaction



14

Configural Equivalence

Differences in conceptualization of job satisfaction

- Different factor structures of job satisfaction: Singaporeans view co-workers as a part of the nature of their work, Americans perceive co-workers as being related to supervisors
  - Spector & Wimalasiri (1986) *Int'l Review of Applied Psyc*
- Education level → Cognitive complexity → Dimensions of pay satisfaction
  - Carraher & Buckley (1996) *JAP*
- Egyptian – job security is taken for granted because Egypt restricts the ability of organizations to terminate employment
  - Parnell & Hatem (1997) *Int'l J of Value-Based Mgt*

15

Metric Equivalence

Differences in strength of relationship between a particular belief (item) and its underlying dimension

- Factor loadings of *independent thought* and *challenge* are lower for the Egyptian sample than for the Western Sample
  - Parnell & Hatem (1997) *Int'l J of Value-Based Mgt*
- People in one culture may be more sensitive to differences in a scale item than people from other cultures

Scalar Equivalence

Differences in response threshold

- Different standards/expectations of satisfaction/dissatisfaction

16

Uniqueness Equivalence

Differences in familiarity with a particular item

- Differences in uniqueness variance between the US and Australian samples on job satisfaction
  - Ryan, Chan, Polyhart, & Slade (1999) *PPsyc*

Construct Variability Equivalence

Differences in strength of culture

- Existence of within culture variation or sub-culture

17

Construct Relations Equivalence

Differences in construct relations

- National wealth, national social security, cultural individualism, and cultural power distance moderate the relationship between intrinsic job characteristics and job satisfaction
  - Huang & Van de Vliert (2003) *J of OB*

Differences in factor loadings of Second-Order Constructs

- Factor loadings of pay satisfaction on overall satisfaction are lower among Egyptian managers than Western managers
  - Parnell & Hatem (1997) *Int'l J of Value-Based Mgt*

18

Latent Mean Equivalence

Differences in level of Constructs

- Americans are more satisfied with their jobs than the Japanese
  - Lincoln & Kalleberg (1985) *American Sociological Review*
- Academics in the US are the most satisfied in 8 countries
  - Lacy (1997) *Int'l J of Higher Edu and Edu Planning*

19

Challenges

How to differentiate measurement artifacts from theoretical predictions

- *Theoretical explanation – Operationalize the cause and to examine whether the lack of invariance is due to it and nothing else*
- *Triangulation: Identify another scale that measurement invariance exists*

Develop testable propositions about the specific effects of cultures/values/norms and levels of economic development as they relate to measurement of constructs in a broad sense

- *Do collectivists systematically differ from individualists in how they view constructs central to organizational theories?*
- *Do subjects from high context cultures and those from low context cultures view constructs or use scales differently?*

20

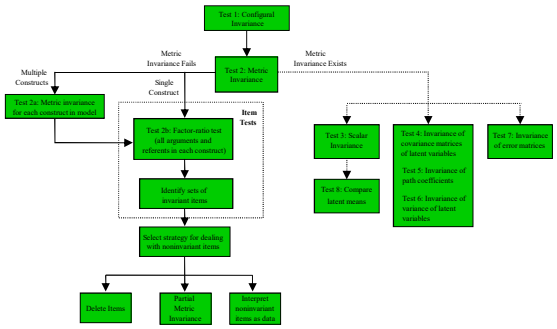
Summary

- Measurement (Non-)Equivalence is not necessarily measurement artifacts
- Whether non-equivalence is unintentional or is predicted on a theoretical basis
- Should be more careful in instrument development and research design

Interpret Non-Equivalence

21

Measurement Equivalence Tests



22

A Direct Comparison Approach for Testing Measurement Invariance

Cheung & Lau (2012). A Direct Comparison Approach for Testing Measurement Invariance. Organizational Research Methods, 15, 167-198

23

Background

- Before making meaningful comparisons across groups in social sciences, researchers need to identify the survey items that fail measurement equivalence/invariance (ME/I)
- Common methods for testing ME/I
  - Likelihood ratio test (LRT; Bollen, 1989)
  - ΔCFI (Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008)
  - Modification index (Marsh & Hocevar, 1985; Yoon & Millsap, 2007)

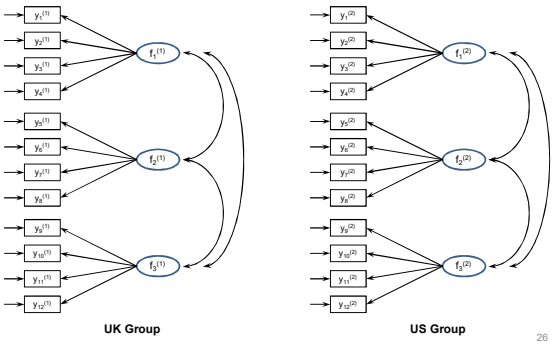
24

Background

- Purpose:
  - To illustrate an M*plus* procedure to estimate the BC bootstrap confidence intervals for testing ME/I, an extension of the procedure for testing mediation effects (Lau & Cheung, 2012)

25

Hypothetical Model for Testing ME/I



26

M*plus* Program for Testing Metric Invariance: ME1.inp

```
TITLE: Example of Bootstrapping CI for Metric Invariance in 2 Samples
DATA: FILE is UKUS.DAT;
VARIABLE: NAMES ARE y1-y12 g;
          GROUPING IS g (3 = UK 4 = US);
MODEL:    f1 BY y1@1 y2 (LX21A)
                y3 (LX31A)
                y4 (LX41A);
          f2 BY y5@1 y6 (LX62A)
                y7 (LX72A)
                y8 (LX82A);
          f3 BY y9@1 y10 (LX103A)
                y11 (LX113A)
                y12 (LX123A);
          [y2] (TAU2A);
          [y3] (TAU3A);
          [y4] (TAU4A);
          [y6] (TAU6A);
          [y7] (TAU7A);
          [y8] (TAU8A);
          [y10] (TAU10A);
          [y11] (TAU11A);
          [y12] (TAU12A);
```

The **MODEL** command describes the overall measurement model to be estimated for each group.

In this model, y1, y5, and y9 are chosen to be the referents and constrained to unity.

The labels of the parameters are assigned in brackets such that each ends with the letter "A", e.g. LX21A and TAU2A.

27

```
MODEL US: f1 BY y2 (LX21B)
              y3 (LX31B)
              y4 (LX41B);
          f2 BY y6 (LX62B)
              y7 (LX72B)
              y8 (LX82B);
          f3 BY y10 (LX103B)
              y11 (LX113B)
              y12 (LX123B);
          [y2] (TAU2B);
          [y3] (TAU3B);
          [y4] (TAU4B);
          [y6] (TAU6B);
          [y7] (TAU7B);
          [y8] (TAU8B);
          [y10] (TAU10B);
          [y11] (TAU11B);
          [y12] (TAU12B);
```

The **MODEL US** command describes how the measurement model of the US group differs from the overall model (i.e. the model of the UK group).

Specifically, the label of each parameter in the US group is different from that in the UK group such that each ends with letter "B", e.g. LX21B and TAU2B.

28

```
MODEL CONSTRAINT:
    NEW (LX1D21 LX1D31 LX1D41);
    LX1D21 = LX21A - LX21B;
    LX1D31 = LX31A - LX31B;
    LX1D41 = LX41A - LX41B;
MODEL CONSTRAINT:
    NEW (LX5D62 LX5D72 LX5D82);
    LX5D62 = LX62A - LX62B;
    LX5D72 = LX72A - LX72B;
    LX5D82 = LX82A - LX82B;
MODEL CONSTRAINT:
    NEW (LX9D103 LX9D113 LX9D123);
    LX9D103 = LX103A - LX103B;
    LX9D113 = LX113A - LX113B;
    LX9D123 = LX123A - LX123B;
ANALYSIS: BOOTSTRAP = 1000;
OUTPUT: CINTERVAL(BCBOOTSTRAP);
```

The **MODEL CONSTRAINT** command and the **NEW** option allow the creation of new parameters (i.e. LX1D21 to LX3D41), e.g. the new parameter LX1D21 is defined as the difference between LX21A and LX21B.

29

Testing Metric Invariance with M*plus* – Output

CONFIDENCE INTERVALS OF MODEL RESULTS

	Lower 0.5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper 0.5%
New/Additional Parameters							
LX1D21	-0.761	-0.665	-0.600	-0.281	0.045	0.123	0.220
LX1D31	-0.930	-0.742	-0.652	-0.197	0.265	0.363	0.525
LX1D41	-0.667	-0.519	-0.431	-0.126	0.193	0.260	0.380
LX5D62	-0.397	-0.307	-0.265	-0.109	0.071	0.094	0.172
LX5D72	-0.182	-0.138	-0.102	0.055	0.217	0.250	0.320
LX5D82	-0.112	-0.073	-0.054	0.081	0.227	0.255	0.318
LX9D103	-0.641	-0.541	-0.479	-0.239	0.011	0.077	0.195
LX9D113	-0.620	-0.477	-0.429	-0.161	0.125	0.191	0.301
LX9D123	-0.530	-0.407	-0.360	-0.103	0.129	0.189	0.285

30

Testing Scalar Invariance with M*plus* – Output

CONFIDENCE INTERVALS OF MODEL RESULTS

	Lower 0.5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper 0.5%
New/Additional Parameters							
TAU1D2	**0.135**	0.191	0.216	**0.340**	0.489	0.514	**0.559**
TAU1D3	**0.069**	0.125	0.163	**0.325**	0.511	0.559	**0.650**
TAU1D4	-0.295	-0.249	-0.226	-0.102	0.046	0.077	0.130
TAU5D6	-0.137	-0.104	-0.086	-0.006	0.056	0.071	0.098
TAU5D7	**-0.341**	-0.312	-0.295	**-0.217**	-0.147	-0.134	**-0.102**
TAU5D8	-0.109	-0.054	-0.035	0.048	0.114	0.125	0.149
TAU9D10	-0.346	-0.292	-0.273	-0.131	0.001	0.022	0.073
TAU9D11	-0.181	-0.100	-0.080	0.070	0.220	0.246	0.288
TAU9D12	-0.290	-0.235	-0.211	-0.071	0.066	0.089	0.143

31

Comparing Latent Means with M*plus* - Output

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Group UK				
Means				
f1	0.000	0.000	999.000	999.000
f2	0.000	0.000	999.000	999.000
f3	0.000	0.000	999.000	999.000
Group US				
Means				
f1	-0.400	0.049	-8.201	0.000
f2	0.019	0.046	0.411	0.681
f3	-0.053	0.037	-1.413	0.158

CONFIDENCE INTERVALS OF MODEL RESULTS

	Lower 0.5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper 0.5%
Group US							
Means							
f1	**-0.525**	-0.493	-0.478	**-0.400**	-0.318	-0.303	**-0.271**
f2	-0.090	-0.064	-0.051	0.019	0.100	0.112	0.141
f3	-0.148	-0.129	-0.117	-0.053	0.005	0.019	0.042

32

Discussion

- The BC confidence interval procedures:
  - Give an estimate of the difference between 2 parameters across groups and a confidence interval for the difference
  - Correct the bias in the bootstrapped sampling distribution
  - Allow all item-level tests for all constructs in one model estimation
  - Allow factor-ratio tests in one model estimation

33

Additional References

- Cheung, G. W. & Lau, R. S. (2012). A direct comparison approach for testing measurement invariance. *Organizational Research Methods, 15*, 167-198.
- Cheung, G. W., & Roger B. Rensvold. (2002). "Evaluating Goodness-of-Fit Indices for Testing Measurement Invariance", *Structural Equation Modeling Journal, 9(2),* 233-255.
- Cheung, G. W. & Roger B. Rensvold. (2000) "Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research using Structural Equations Modeling", *Journal of Cross-Cultural Psychology, 31(2)*, 187-212.
- Cheung, G. W. (1999) "Multifaceted Conceptions of Self-Other Ratings Disagreement", *Personnel Psychology, 52(1),* 1-36.
- Cheung, G. W., & Roger B. Rensvold (1999). "Testing Factorial Invariance Across Groups: A Reconceptualization and Proposed New Method", *Journal of Management, 25(1)*, 1-27.

34

Additional References

- Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance: Methods and procedures. *Organizational Research Methods*, *5*, 139-158.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4-70.

35