

Luye Chang, Ph.D., Select International, Inc.
 Janet L. Barnes-Farrell, Ph.D., & Dev K. Dalal, Ph.D. University of Connecticut

INTRODUCTION

Despite the pervasive international use of 360-degree leadership ratings, there is a paucity of empirical validation evidence to support this practice. This study addressed three research questions:

Q1. Are measurements equivalent across organizational levels in 360-degree ratings?

- Basic assumption of 360-degree ratings: assessors from different organizational levels interpret and respond to leadership scales similarly.
- *Conceptual disagreement*: raters could conceptualize definitions of performance/leadership dimensions or the relative importance of observed behaviors differently (Borman, 1997)
- *Psychometric disagreement*: differences in response style (Cheung, 1999).
- A handful of studies have examined cross-rater measurement equivalence, but only on North American samples.
- Part 1 of the study: examined Chinese sample as a demonstration of the importance of cross-rater measurement equivalence studies of 360-ratings in other countries

Q2. Are measurements equivalent between U.S. and China?

- Before meaningful comparisons between managers from different countries can be made, whether or not ratings are equivalent across countries must be addressed.
- Part 2 of this study: inspected measurement equivalence between U.S. and China between multiple rating sources, including self-raters, supervisors, peers, and subordinates.

Q3. Are there differences in self-other agreement between the U.S. and China?

- *Self-Other Agreement (S-OA)*: discrepancies between an individual's self-ratings and other-ratings are not measurement errors, but meaningful indicators of an individual's *self-awareness* (Cheung, 1999; Church, 1997).
- Cross-national differences in S-OA are theoretically plausible yet seldom empirically investigated.
- Part 3 of the study: examined the extent to which self-other agreement in 360-degree ratings differ by country—the United States and China.
 - Power Distance (acceptance to power discrepancy)
 - Individualism (tendency to be self-reliant)
 - Assertiveness (tendency to be confrontational in interpersonal interaction)
 - Uncertainty Avoidance (resistance to unpredictable future and ambiguity)

METHOD

Sample. Participants in this study were leaders from the United States and China. In the 360-degree feedback program, they rated themselves – using a proprietary instrument – on a number of leadership competencies and were evaluated on the same competencies by their supervisors, peer colleagues, and subordinates. Sample sizes varied across competencies from 1,389 to 2,622 in the U.S. sample and from 226 to 340 in China sample.

Measures. Each competency is measured by a set of behavioral descriptors, ranging from 4 to 6 items. A 5-point Likert scale (1=*not proficient*; 5=*fully proficient*) was used. Four competencies: Team Building, Coaching, Decision Making, and Driving for Results.

Data Analysis. All research questions were examined under the Mean and Covariance Structure (MACS) framework (Ployhart & Oswald, 2004), which assesses variance, covariance, and latent means in a series of nested SEM models. Measurement equivalence testing followed Vandenberg and Lance's (2000) guidelines.

Figure 1. Latent Means of Self-Ratings and Other-Ratings



RESULTS

Q1: Are measurements equivalent across organizational levels in 360-degree ratings? (Table 1)

- Measurement equivalence was tested using MACS given that self-, supervisor-, peer-, and subordinate-ratings were not independent from each other.
- Identical factor form (*configural invariance*) and factor loadings (*metric invariance*) were supported.
- Partial *scalar invariance* (item intercepts) was observed in team building (China), coaching (U.S.), and driving for results (U.S. and China).

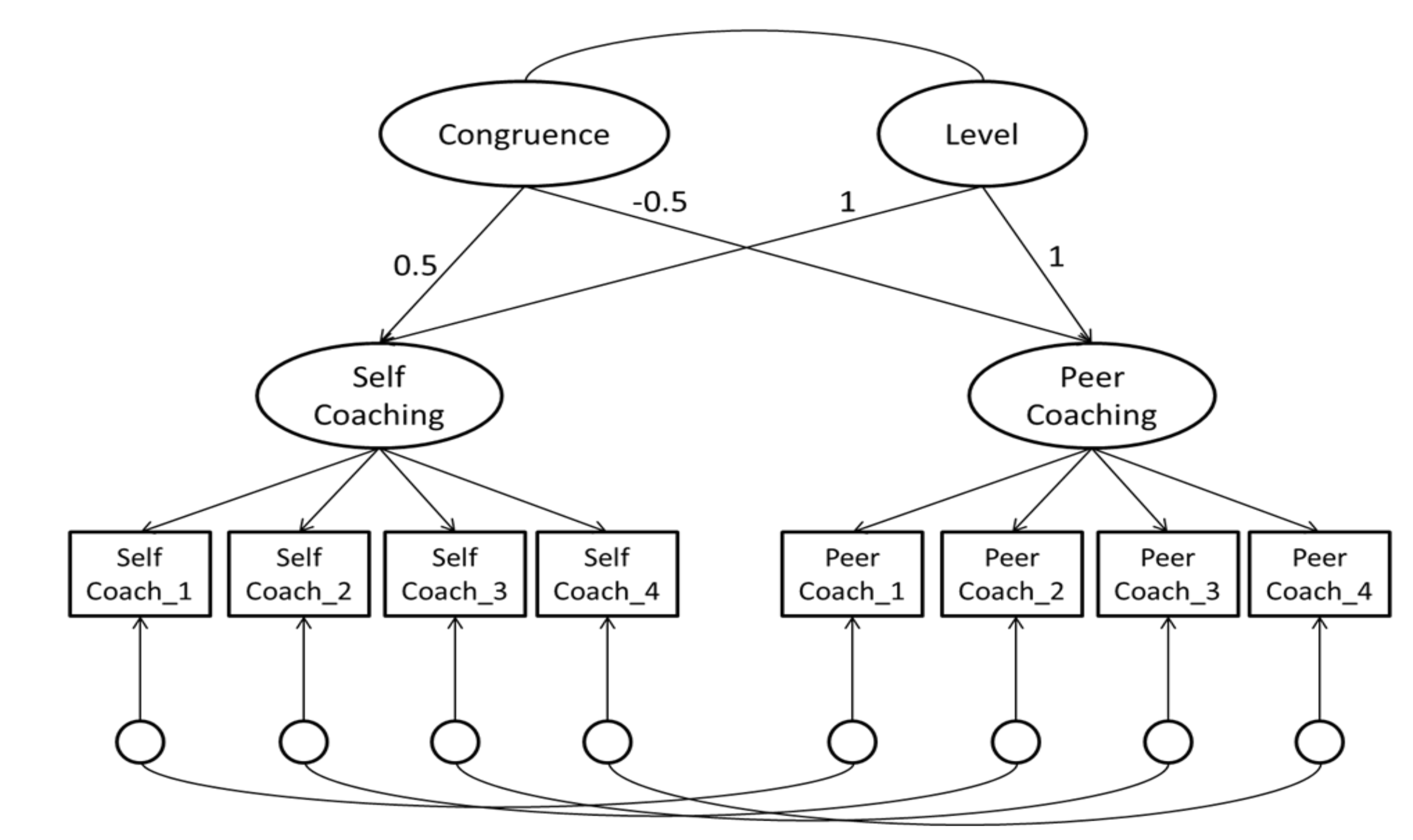
Q2. Are measurements equivalent between U.S. and China?

- Multi-group confirmatory factor analysis (MGCFAs) was performed.
- Factor forms, factor loadings, and item intercepts were found to be equivalent between two countries for all competencies.
- Chinese ratings produced smaller residual variances of items for *team building* and *driving for results*.

Q3. Are there differences in self-other agreement between the U.S. and China? (Figure 1)

- Latent congruence modeling (Cheung, 2009) was examined (an example is shown in Figure 2).
- *Coaching* was the only factor in which self-supervisor discrepancies and self-peer discrepancies differed significantly from the U.S. to China.
- National differences on self-subordinate discrepancies were detected on two competencies—*decision making* and *driving for results*.
- Chinese raters, both self-raters and other-raters, provided markedly lower ratings than their American counterparts.
- Subordinate ratings were relatively equal between the U.S. and China.

Figure 2. Latent Congruence Model of Self-Peer Agreement



DISCUSSION

Theoretical Implications

- Raters across organizational levels perceive and interpret leadership constructs similarly, whether they are from China or the U.S.
- Psychometric equivalence, however, was not as strongly supported as conceptual equivalence. Possible explanations:
 - (1) raters are expected to provide somewhat unique insights from observing leader behaviors that are not necessarily revealed to other raters;
 - (2) disparate motivations drive raters to contaminate true evaluations with differential degree of leniency or stringency bias.
- Raters from different countries share a common conceptualization and response pattern to leadership constructs.
- When measurement error variances were statistically controlled, national differences in self-other discrepancies remained significant in some competencies.

Conclusion

The present study is the first to examine cross-national measurement equivalence of 360-degree leadership ratings, and compare self-other discrepancies on the latent construct level. Although results indicated that 360-degree leadership assessment can be validly delivered in Chinese settings, efforts must be made to check for differential response patterns between countries.

Table 1 Measurement Equivalence Results of Team Building between Self and Other-Ratings

U.S. Sample	Δdf	Δχ ²	CFI	ΔCFI	TLI	RMSEA	RMSEA 90% CI
1. Configural Invariance			.951		.930	.052	.049 - .055
2. Metric Invariance	15	41.34	.949	.002	.932	.051	.048 - .054
3. Scalar Invariance	15	164.00	.940	.009	.925	.054	.051 - .057
4. Uniqueness Variance Invariance	18	130.43	.933	.007	.922	.055	.052 - .058
5. Factor Variance Invariance	3	350.54	.911	.022	.898	.063	.060 - .065
5'. Factor variance partial	1	12.13	.932	.001	.921	.055	.052 - .058

Chinese Sample	Δdf	Δχ ²	CFI	ΔCFI	TLI	RMSEA	RMSEA 90% CI
1. Configural Invariance			.966		.952	.051	.040 - .061
2. Metric Invariance	15	20.42	.965	.001	.953	.050	.040 - .060
3. Scalar Invariance	15	54.38	.954	.011	.942	.056	.046 - .065
3'. Scalar partial	13	32.84	.959	.006	.948	.052	.043 - .062
4. Uniqueness Variance Invariance	18	35.11	.954	.005	.946	.053	.044 - .062
5. Factor Variance Invariance	3	33.37	.946	.008	.937	.058	.049 - .067

Note. Δdf=difference of degree of freedom between this model and the less restricted model above; CFI=Comparative fit index; ΔCFI=difference of CFI between this model and the less restricted model above; TLI=Tucker-Lewis index; RMSEA 90% CI=90% confidence interval of RMSEA. Models highlighted in grey shading yielded significantly different model fit from the comparison model.