

Cross-cultural measurement of supervisor trustworthiness: An assessment of measurement invariance across three cultures

S. Arzu Wasti^{a,*}, Hwee Hoon Tan^{b,1}, Holly H. Brower^{c,2}, Çetin Önder^{d,3}

^a Sabanci University, Faculty of Management, Orhanli 34956 Tuzla, Istanbul, Turkey

^b Singapore Management University Lee Kong Chian School of Business, 50 Stamford Road, Singapore 178899, Singapore

^c Calloway School of Business and Accountancy, Wake Forest University, Reynolda Station, Winston-Salem, NC 27109, USA

^d Sabanci University, Faculty of Management, Orhanli 34956 Tuzla, Istanbul, Turkey

Abstract

This paper contributes to the research on supervisor trustworthiness by assessing the measurement equivalence of the trust scales developed by Mayer and Davis [Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, 84, 123–136] across three samples: U.S., Turkey and Singapore. This study found the trust scale to have poor psychometric properties across the board, rendering invariance tests inappropriate. Analysis of the antecedents of trust scales supported the metric equivalence of the integrity measure, but several items of the ability and benevolence scales appeared to be interpreted differently by respondents from collectivist-high power distant versus individualist-low power distant cultures. We advocate the formation of a multinational team of trust and leadership scholars to develop scales in which items reflect not a single culture but are more applicable both in meaning and choice of expression to many cultures.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Supervisor trustworthiness; Cross-cultural; Measurement invariance

1. Introduction

Research on interpersonal trust, its antecedents and subsequent organizational effects has spanned several decades, but refining and developing our understanding of trust in organizational settings has been particularly strong in the last decade (e.g. Dirks & Ferrin, 2002; Mayer, Davis, & Schoorman, 1995; Rousseau, Sitkin, Burt, & Camerer, 1998). Recent research has focused on trust as a psychological state that develops over time between two or more individuals developed as a result of a social exchange process (Brower, Schoorman, & Tan, 2000; Currall & Judge, 1995; Hosmer, 1995; Kramer & Tyler, 1996; Whitener, Brodt, Korsgaard, & Werner, 1998).

* Corresponding author. Tel.: +90 216 4839662; fax: +90 216 483 9699.

E-mail addresses: awasti@sabanciuniv.edu (S.A. Wasti), hhtan@smu.edu.sg (H.H. Tan), browerhh@wfu.edu (H.H. Brower), cetinonder@su.sabanciuniv.edu (Ç. Önder).

¹ Tel.: +65 68280524; fax: +65 68280777.

² Tel.: +1 336 758 6174.

³ Now at Dokuz Eylül University Faculty of Business, Buca, İzmir. Tel.: +90 232 453 5060; fax: +90 232 453 5062.

Since the publication of a special issue of *Academy of Management Review* in 1998 devoted to trust, the model of the construct developed by Mayer et al. (1995) has become widely accepted and used in organizational research (Dirks, 1999; Rousseau et al., 1998). Trust as defined by Mayer et al. is specific to the domain of interpersonal work relationships, particularly the relationship between supervisors and subordinates. In this model, trust is defined as “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party.” (Mayer et al., 1995: 712). This vulnerability leaves the trustor open to the possibility of disappointment or betrayal. Mayer et al. (1995) further assert that the antecedents to trust include perceptions of the trustee’s (supervisor’s) ability, integrity and benevolence toward the trustor (subordinate) in the situation. Ability is an assessment of the group of skills, competencies, and characteristics that the trustor believes the trustee has in a given context. Benevolence is the degree to which the trustee is believed to have the best interests of the trustor in mind. Finally, integrity is the extent to which the trustor believes the trustee adheres to a set of principles that the trustor finds acceptable. In addition, each trustor has a propensity to trust that is an individual difference that sets the base level of trust at the beginning of a relationship.

1.1. Linkages between leadership and trust

Using this model of trust, Brower et al. (2000) developed a model of relational leadership. Since both trust and leader–member exchange (LMX), a popular model of the leadership relationship between supervisors and subordinates, are shaped through the social exchange process, they integrated these two theories to model trust as a motivational engine for leaders. Trust has been included as one of the foundations of relationship quality between supervisors and their subordinates in LMX research (Graen & Uhl-Bien, 1995; Schriesheim, Neider, & Scandura, 1998). Empirical research on LMX has found that trusting relationships, characterized as high LMX, are associated with increased subordinate job satisfaction (Liden & Maslyn, 1998; Seers, 1989), performance (Liden & Maslyn, 1998; Scandura & Schriesheim, 1994), commitment (Liden & Maslyn, 1998; Green, Anderson, & Shivers, 1996), citizenship behaviors (Anderson & Williams, 1996) and reduced turnover (Liden & Maslyn, 1998). When studied outside of the LMX construct, trust has had similar effects on all of these same outcomes (Brower et al., 2000). Simply put: when leaders are viewed as trustworthy, subordinates are motivated to increase performance, demonstrate satisfaction and stay with the organization.

In other leadership theories that look at distant leaders, trust has also been a significant factor contributing to leader effectiveness. Specifically, theorists have asserted that trust in the leader is critical for charismatic or transformational leaders (House, 1976; Podsakoff, MacKenzie, Moorman, & Fetter, 1990; Tan & Wee, 2002). In fact, trust has been found to moderate the relationship between transformational/transactional leadership styles and subordinate outcomes (Pillai, Schriesheim, & Williams, 1999; Podsakoff et al., 1990). Two cross-cultural studies are pertinent to this discussion. Tan & Wee (2002) studied the National Rally Day speeches of a Singaporean national leader and found that in his speeches, this charismatic leader portrayed self-confidence as a means of building trust from his followers. However, he did not amplify values and mention past success, which would be considered trust building tactics in a Western context. More recently, Casimir, Waldman, Bartram & Yang (2006) found that trust mediated the relationship between transformational leadership and subordinate citizenship behaviors in Australian subjects but not in Chinese subjects. They argued that in collectivist cultures heavily influenced by Confucian values supportive of power distance, individuals may be more accepting of autocratic leadership practices, but acceptance does not necessarily translate into trust in the leader. Further, they proposed that in low power distant cultures, transformational leadership behaviors such as showing respect and consideration, articulating a vision shared by followers, and showing determination toward achieving goals engender feelings of trustworthiness.

Both of these studies demonstrate the importance of understanding how trust develops across different contexts because leadership styles may have very different impacts on subordinate outcomes. Indeed, Doney, Cannon & Mullen (1998) underscored the importance of understanding the influence of culture on trust development by drawing attention to an increasingly multicultural work environment. Our premise is that to better understand the cultural implications of trust in organizational settings, we need measures that are valid across cultural boundaries. While the model developed by Mayer et al. (1995) has received both theoretical (Dirks, 1999; Dirks & Ferrin, 2002; Rousseau et al., 1998) as well as some empirical support (Davis, Schoorman, Mayer, & Tan, 2000; Mayer & Davis, 1999; Mayer & Gavin, 2005), the validity evidence remains limited to the U.S. context and the cross-cultural implications of this model have not been thoroughly considered. Consequently, we cannot make assertions about the explanatory power of trusting in leaders in other cultures.

1.2. Cultural influences on measures of trustworthiness and trust

In the present paper, we propose that the theoretical model by Mayer et al. (1995) is valid for understanding trust between supervisors and subordinates across cultural contexts. However, we hold that culture may reflect itself in the operationalization of the constructs, in the mean levels of propensity to trust, in the relative importance of the factors of trust in predicting trust, and finally, in the emergence of culture-specific antecedents and outcomes of trust. This contention, while implicit in the original model, is supported by other research. For instance, in a qualitative study of interpersonal trust in a Confucian influenced setting, Tan & Chee (2005) observed that ability, benevolence and integrity indeed emerged as antecedents of trust, along with various culture-specific antecedents such as filial piety and perseverance. Also, cultural values appeared to reflect themselves in the relative importance of the various antecedents. Likewise, in their model of national culture and development of trust, Doney et al. (1998) argued for the universal relevance of the trust-building processes proposed by Mayer et al. (1995) but theorized that the prevalence of a given process would be a function of cultural values.

The focus of this paper is to test for metric equivalence in the operationalization of trust and its antecedents for the Mayer et al. (1995) model. Over the past half decade, cross-cultural researchers have heightened the awareness of establishing measurement equivalence across culturally diverse groups (e.g. Van de Vijver & Leung, 1997; Vandenberg & Lance, 2000). Establishment of metric invariance indicates that individuals from different cultures respond to items in the same way. Comparisons between cultural groups are appropriate only after ascertaining measurement equivalence. Thus, establishing the metric equivalence of the existing scales of factors of trust (FOT, also known as the antecedents of trust) and trust will constitute a timely and necessary first step (Steenkamp & Baumgartner, 1998). Should metric equivalence be supported, we would then be able to pursue substantive cross-cultural hypotheses regarding trust formation between supervisors and their subordinates.

The countries represented in the current study are the U. S., Singapore and Turkey. Singapore and Turkey provide an informative cultural contrast to the U.S. and can be argued to constitute a stringent test of generalizability. Specifically, the U.S. has repeatedly emerged as one of the most individualistic societies in the world, whereas Turkey and Singapore cluster as more collectivist cultures (Hofstede, 1980; Oyserman, Coon, & Kemmelmeier, 2002). The essential difference between individualism and collectivism is with respect to the construal of the self (Markus & Kitayama, 1991; Triandis, 1995). In individualism, the self is viewed as independent, meaning that individuals seek to maintain their independence from others by giving priority to their own thoughts, needs and feelings and by focusing on their unique attributes. In collectivism, the construal of the self is interdependent. In other words, the most meaningful social units in collectivist societies are the groups to which people belong, such as the family, and one's identity is defined by membership in these groups. In collectivistic societies, the desire to maintain ingroup harmony results in a tendency to subordinate personal preferences and priorities to those of the group (Triandis, 1995). In addition, while the U.S. emerges as relatively low on power distance, Singapore and Turkey are characterized as high power distance cultures (Hofstede, 1980; Kabasakal & Bodur, 2002). Power distance reflects the emphasis given to hierarchical relations in a society. In a culture with large power distance, there is stronger adherence to authoritarian norms as well as greater dependence on and acceptance of authority. On the other hand, low power distance cultures are characterized by smaller inequality, less concentration of power, more egalitarian relationships and lower levels of conformity.

The cultural dimensions of individualism–collectivism and power distance have received considerable theoretical attention (e.g., Triandis, 1995; Triandis & Gelfand, 1998) and empirical research has consistently linked these dimensions to differential implications for workplace behavior and work-related attitudes (see for example, Smith, Fischer & Sale, 2001 for a review of this literature), including leadership (House et al., 2004). Given that these cultural dimensions remarkably influence the way people understand and interpret their context, including the work context, we argue that they will have implications for the construction and the interpretation of survey instruments used in cross-cultural organizational research (Riordan & Vandenberg, 1994). In the following paragraphs, we elaborate on the manifestations of certain cultural assumptions inherent in the items that constitute the Mayer et al. (1995) scales and whether it is likely that the scales will elicit the same conceptual frames of reference across cultures.

With respect to the FOT scales, we argue that the operationalization of the ability scale items may be particularly culture-specific. In individualistic cultures where the emphasis is on the individual's internal attributes such as traits, abilities and motivation (Markus & Kitayama, 1991), a person's identity is largely based on personal accomplishments (Doney et al., 1998). In such contexts, individual task performance is recognized and rewarded whereas in collectivistic

cultures, relationships are more highly valued than individual achievement (Hofstede, 1980). Research further indicates that collectivistic cultures consider competence to be secondary to attributes of the person such as family background or social class, and the latter are viewed to be more informative qualifications in terms of assessing trustworthiness, for instance, in hiring decisions (e.g., Adler & Jelinek, 1986). Therefore, we anticipate that items measuring ability by explicitly tapping task performance may not evoke the same conceptual frames across individualistic and collectivistic cultures. We also expect that variation in power distance may have implications for the metric equivalence of this scale. Specifically, in high power distance countries subordinates may not be familiar with assessing the competence of people in authority, and consequently, they may not read these items similarly. For instance, they may naturally assume that their supervisor is competent and may feel that evaluating the competence of a supervisor may come across as insubordination (Riordan & Vandenberg, 1994).

On the other hand, the benevolence scale, which is composed of items reflective of attachment, relatedness and a non-instrumental concern for ingroup welfare can be expected to evoke similar conceptual frames for collectivists as these attributes exhibit the core element of collectivism, namely, interdependence. Furthermore, these items also are congruent with the notion of paternalism, which is characteristic of high power distant and collectivist cultures. Paternalism is described as a subordinate-superior relationship whereby those in authority assume the role of a parent and consider it an obligation to provide support and protection to those under their care (Aycan, 2001). In paternalistic cultures, managers are concerned with and interested in personal as well as professional welfare of their subordinates. Cross-cultural research has characterized both Turkey and Singapore as paternalistic in their managerial style (Aycan et al., 2000; Putti & Chong, 1987). Thus, we expect respondents from Turkey and Singapore to similarly understand questions assessing a supervisor's benevolence.

Albeit for different reasons, invariance can be predicted for the integrity scale, which taps whether the trustee follows through on expectations or if the individual can be counted on that his or her words and actions are congruent. The items that constitute this scale can be argued to manifest a generalizable notion of consistency and credibility and do not seem to be culture-bound with respect to representing different self-construals or generating hierarchical frames of reference. In other words, the current scale appears to be representing a culturally neutral operationalization of the integrity construct. In contrast, respondents from different cultural backgrounds can be expected to react differently to the specific items that constitute the trust scale. We argue that the current trust scale in effect signals a concern for individual welfare and a notion of personal fate, which appear to be more congruent with the defining attribute of individualism, namely, the assumption that individuals are independent of one another. The items, which imply a contractual and calculating relationship between the supervisor and the subordinate, are not particularly fitting with the paternalistic attributes of the Turkish and the Singaporean culture, either. Thus, it is likely that respondents from the three cultures will not use a common frame of reference when responding to the items of the trust scale.

2. Method

2.1. US sample

Data from the US sample was collected from a restaurant chain, which has approximately 800 employees. Data was collected personally during work hours. All participants were given a cover letter stating the purpose of this research with an assurance of confidentiality. A total of 334 usable questionnaires were collected (response rate of 41.8%). There was about an equal proportion of men and women (50.4% and 48.6% respectively) and 54.0% of the participants were 21–30 years old. About half (48.8%) of the respondents reported at least a high school diploma and some college education while 36.7% had a college degree. Half of the respondents (51.3%) had been with the restaurant chain for less than a year.

2.2. Singapore sample

247 questionnaires were distributed to 4 organizations (a large government organization, a local media organization, an army company, and a non-profit organization) in Singapore. Data was collected personally from each individual although they were allowed to take the surveys home. All participants were given a cover letter stating the purpose of this research with a confidentiality assurance. As English is the accepted business language, the English questionnaire was used in Singapore. 215 people responded, yielding a response rate of 87%. Of the 215 obtained, 207 questionnaires (final response rate of 83.8%) were error free. 92.3% of the respondents had been in the organization for less than

6 years. Seventy-four percent of the respondents were male and 78.7% of them were between the ages of 18–25 years. All respondents were at least college graduates.

2.3. Turkish sample

Data were collected from 434 employees in 6 different branches of an automotive retailer (74% return rate). This sample consisted mostly of male employees (81%). Approximately 74% of the participants had at least a high school degree and 22% had at least a college degree. Surveys were conducted on company time, either in training rooms or in catering areas. In these sessions, the researcher explained the purpose of the study, assured anonymity and personally collected survey forms. Data were collected via a paper-and-pencil questionnaire in Turkish. One bilingual academic translated the questionnaire into Turkish and the Turkish version was subsequently back-translated to English by another bilingual academic. The two translators then met to finalize the Turkish version of the instrument.

In order to ascertain that subsequent findings were not biased by demographic differences between the samples, we first compared the samples on three demographic variables: gender, education and organizational tenure. Chi-square analyses of gender and education variables and a univariate ANOVA of the tenure variable revealed significant differences between three samples. Thus, in the next step, we investigated whether ability, benevolence, integrity and trust scores were affected by demographic differences using the procedure outlined by Riordan & Vandenberg (1994). First, two-way ANOVAs were run on each scale for gender and education. The analyses revealed that, except for a single interaction (country-gender interaction for the trust score) neither the main effects nor the interactions were significant for any of the four variables of interest. Second, ability, benevolence, integrity and trust scores were regressed on tenure, country, and tenure–country interaction. The results showed that the main effect of tenure and the effect of tenure–country interaction were not significant. Overall, we concluded that ability, benevolence, integrity and trust scores were not biased by the demographic differences between the samples.

2.4. Measures

We used the measures of the FOT and trust that were originally developed by Schoorman, Mayer & Davis (1996) and published in Mayer & Davis (1999). The latter study, while failing to obtain satisfactory internal consistency, demonstrated acceptable test–retest reliability and discriminant validity for the trust scale. The scales measuring the antecedents of trust, on the other hand, demonstrated satisfactory reliability and validity (e.g., Davis et al., 2000; Mayer & Davis, 1999). The propensity scale, which consistently indicated poor psychometric properties, was not included in this analysis. Accordingly, subordinates were asked to report perceptions of their supervisor's ability (6 items), benevolence (5 items), and integrity (6 items) and to report their own trust in their supervisor (4 items).⁴ All scales used a 5-point Likert-type response with anchors ranging from (1) disagree strongly to (5) agree strongly.⁵ The English versions of the scales are presented in the Appendix.

2.5. Overview of analysis

Prior to testing measurement equivalence, item and scale analyses were carried out to assess the reliability of the scales in each of the samples. Next, we evaluated factorial invariance across the three groups. We first fitted a single-factor specification (Model 1) to the data, which served as the baseline. Then, a two-factor specification (Model 2) that distinguished between trust items and items of the FOT scales was fitted to the data. Model 3 distinguished between trust, ability and a third factor comprising benevolence and integrity items. Finally, Model 4 distinguished between trust,

⁴ We eliminated one benevolence item as there was a slight variation in wording across the different administrations. The analysis reported is based on the remaining four items.

⁵ In the U.S. sample, all scales used a 5-point Likert-type response with anchors ranging from (1) disagree strongly to (5) agree strongly. In the Turkish and Singapore samples, all scales used a 7-point Likert-type response ranging from (1) disagree strongly to (7) agree strongly. We report the analyses based on the original scales in the U.S. sample and rescaled responses for the Turkish and Singapore samples. For the latter samples, the responses were rescaled using the formula (new 5-point score = ((old 7-point score) – 1) * 2/3 + 1). We replicated the analyses with the original 7-point scales for the Turkish and Singapore samples and rescaled responses for the U.S. sample (where the response option 2 was coded as 2.5 and the response option 3 was coded as 5.5, with the extreme anchors and the neutral point coded accordingly). All analyses, irrespective of the coding, yielded the same results.

ability, benevolence and integrity. Though Model 2 and Model 3 displayed significantly better fit to the data, we ran into estimation problems while fitting Model 4, due to collinearity between the FOT scales in the Turkish sample. Therefore, following the procedure outlined in Riordan & Vandenberg (1994) and considering the test of discriminant-convergent validity as premature at this stage of the investigation, we conducted metric invariance analyses separately for each scale.

The invariance analysis, namely Jöreskog’s (1971) Simultaneous Factor Analysis in Several Populations (SIFASP), demonstrates the extent to which the scaling and representativeness of the indicators are similar across samples. More specifically, four constrained models (one for each scale) were estimated simultaneously across the three samples with an equality constraint imposed on the factor loadings. These models were then tested against their respective baselines, i.e., the unconstrained models. In order to assess the effect of the equality constraint, the chi-square statistic for each constrained model was compared to the chi-square statistic for the relevant unconstrained model. In addition, three other fit indices, namely the comparative fit index (CFI; Bentler, 1990), Tucker–Lewis index (TLI) and root mean squared error of approximation (RMSEA) were inspected. An insignificant change in chi-square statistic ($\Delta\chi^2$) was considered evidence for metric invariance given the CFI, TLI and RMSEA displayed an acceptable overall model fit.

The second stage of the analysis involved examination of partial invariance of the scales for which metric invariance could not be established. Investigations of partial invariance are based on the idea that although a full scale may not display invariance across samples some of the scale’s items may be invariant (Vandenberg & Lance, 2000). If only a minority of items is found to be non-invariant, researchers can proceed to investigate substantive questions of interest with scales composed of invariant items (Cheung & Rensvold, 1999). To discover the invariant items, the factor–ratio test was used (Cheung & Rensvold, 1999). The factor–ratio test is superior to alternative procedures in that it eliminates risks associated with inadvertent selection of the referent item. Referent items are the items whose loadings are constrained to be 1 for identification purposes. Unknowingly selecting a non-invariant item as referent results in inaccurate parameter estimates and biased model comparisons (Cheung & Rensvold, 1999; Vandenberg, 2002). The factor–ratio test basically involves constraining the factor loading of a single item one at a time and comparing the chi-square statistic for the constrained model to the chi-square statistic for the unconstrained model. The procedure results in identification of invariant sets, which are items exhibiting invariance when all other members of the set are used as the referent. Fig. 1 presents a hypothetical procedure. The letter entries in cells in Fig. 1a and b denote chi-square change statistics. Each row displays the change in the chi-square statistic due to constraining an item by sequentially using all other items of the scale as the referent. Fig. 1a shows the chi-square change statistics associated with all combinations of referent and constrained items. Fig. 1b shows the insignificant chi-square change statistics only. The upper diagonal entries are eliminated in Fig. 1b because they are redundant (Cheung & Rensvold, 1999). As Fig. 1b

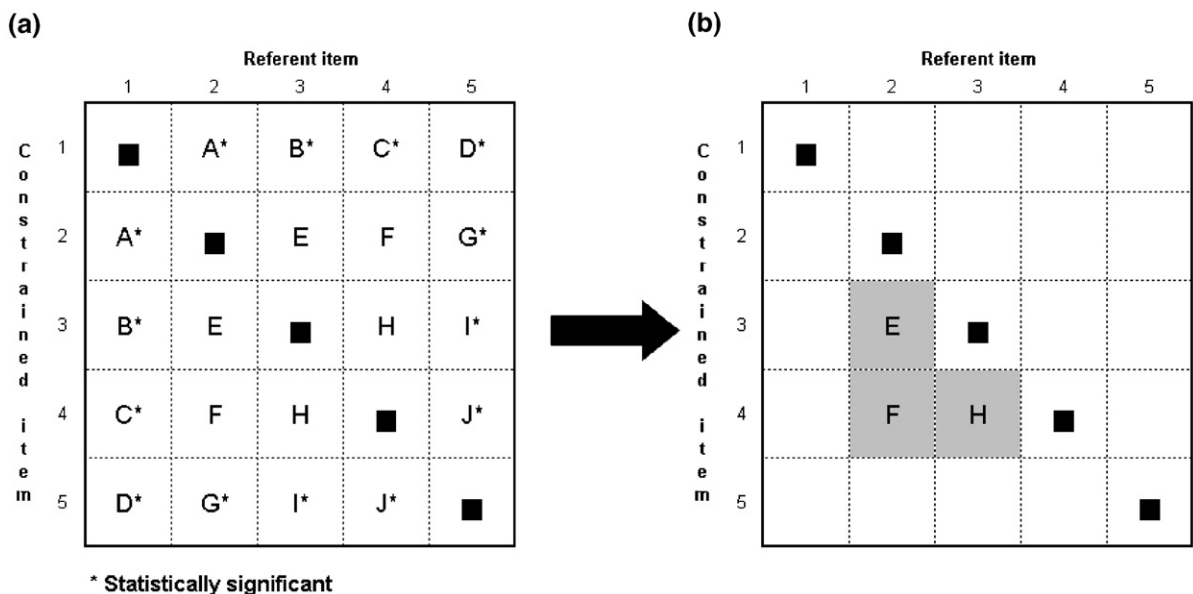


Fig. 1. Factor–ratio test.

Table 1
Interscale correlations, means, standard deviations and internal consistency estimates for the scales

	Turkey						Singapore						U.S.					
	μ	<i>SD</i>	1	2	3	4	μ	<i>SD</i>	1	2	3	4	μ	<i>SD</i>	1	2	3	4
1 Ability	2.54	1.09	.91															
2 Benevolence	2.70	1.16	.83	.85														
3 Integrity	2.62	1.01	.86	.85	.84													
4 Trust	2.95	.94	.61	.64	.66	.55												
1 Ability							3.40	.90	.93									
2 Benevolence							3.13	1.06	.74	.92								
3 Integrity							3.10	.91	.69	.86	.88							
4 Trust							2.82	.80	.53	.64	.71	.68						
1 Ability													3.94	.77	.94			
2 Benevolence													3.60	.91	.76	.92		
3 Integrity													3.60	.79	.79	.81	.87	
4 Trust													3.25	.78	.63	.71	.67	.70

All correlations are significant at $p < .001$.

Cronbach alpha estimates of internal consistency are indicated on the diagonals.

shows constraining item 3 by using item 2 as the referent results in an insignificant chi-square change. The same is true for combinations that involve items 2 and 4 and items 3 and 4. Items 2, 3 and 4 thus constitute an invariant set. The results of the measurement invariance analysis are discussed in detail below.

3. Results

3.1. Reliability analysis

Table 1 shows the reliability estimates of the four scales as well as interscale correlations, means, and standard deviations for each of the three samples. Consistent with prior findings with U.S. samples, we observed no reliability problems with the scales assessing the FOTs (ability, benevolence, integrity). The reliabilities for these scales ranged from .84 to .94 across the three samples. However, the trust scale was found to have relatively lower reliability, especially for the non-US samples. In particular, item 3 lowered the alpha estimate considerably in the Turkish sample (when dropped, the alpha increased to .61). Nevertheless, we retained the original version of the scale for the remainder of the analysis.

3.2. Measurement invariance

In the next step of the analysis, the scales were evaluated for metric invariance across the three groups as presented in Table 2. While the results indicated acceptable overall model fit for the ability, benevolence and integrity scales, the

Table 2
Fit indices for the metric invariance assessment

Construct	Model	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	CFI	TLI	RMSEA
Ability	Unconstrained model	172.389	27			.968	.946	.074
	Fully constrained model	212.206	37	39.817 *	10	.961	.953	.070
Benevolence	Unconstrained model	10.476	6			.998	.995	.028
	Fully constrained model	30.035	12	19.559 *	6	.993	.989	.039
Integrity	Unconstrained model	60.483	27			.988	.980	.036
	Fully constrained model	81.174	37	20.691 *	10	.984	.981	.035
Trust	Unconstrained model	103.189	6			.854	.561	.129
	Fully constrained model	152.262	12	49.073 *	6	.789	.683	.110

Note: CFI=Comparative fit index; TLI=Tucker–Lewis Index; RMSEA=Root mean squared error of approximation.

* $p < .05$.

Table 3
Result of factor–ratio test for ability (A1... A6)

Constrained item	Referent					
	A1	A2	A3	A4	A5	A6
A1	1					
A2	10.695	1				
A3	18.045	.825	1			
A4	4.347	2.315	6.704	1		
A5	9.984	7.107	8.551	9.258	1	
A6	13.089	8.094	9.756	11.680	.062	1

Non-shaded entries are significant at $p < .01$ ($df=2$).

trust scale poorly fitted the data. Specifically, for both the unconstrained and the constrained models, the CFI and TLI were below .90 and the RMSEA was well above .08, all of which are in unacceptable ranges (Vandenberg & Lance, 2000). Because plausibility of measurement invariance depends on satisfactory overall model fit (Little, 1997), the trust scale was eliminated from further analysis.

Among the scales that displayed acceptable overall model fit, the equality constraint placed on the factor loadings of the ability scale resulted in a statistically significant decrement in model fit ($\Delta\chi^2(10)=39.817, p < .05$). This finding, as hypothesized, suggests that respondents from different cultural samples may be using different conceptual frames of reference when responding to at least some items of the ability scale. However, contrary to expectations, a similar observation was made for the benevolence scale ($\Delta\chi^2(6)=19.559, p < .05$) and the integrity scale ($\Delta\chi^2(10)=20.691, p < .05$). Therefore, we conclude that equivalence could not be assumed for any of the scales across the Singaporean, Turkish and American employees. Our analysis therefore proceeded to explore partial invariance of the three scales.

Following Cheung & Rensvold (1999), we used the factor–ratio test to identify non-invariant items of the FOT scales. The entries in Tables 3, 4 and 5 are the chi-square change statistics and the shaded cells denote the insignificant ones. As seen in Table 3, of the six ability items, A2, A3, and A4 constituted an invariant set. Using the so-called triangle heuristic proposed by Cheung & Rensvold (1999), two more invariant sets for the ability scale were identified (A2, A3, A5; and A2, A5, A6). As the current literature does not provide any guidance as to what constitutes an appropriate criterion for selection among two or more invariant sets with equal number of items, we proceeded with the invariant item set which yielded the highest internal consistency across the three samples (A2, A3, A4; which yielded an alpha value of .89–.90 across the three samples).

As Table 4 shows, two three-item invariant sets (B1, B2, B3; and B1, B2, B4) were identified for the benevolence scale. Using the same decision rule above, the invariant set that yielded the highest internal consistency across the three samples was selected for further analysis (B1, B2, B3; which yielded an alpha value of .84 to .90 across the three samples). Interestingly, however, the factor–ratio test conducted on the integrity scale revealed that all items were invariant (Table 5). Based on this finding, we consider the earlier results regarding the lack of invariance of the integrity scale to be a statistical artifact, which arguably is not uncommon with the chi-square statistic (Cheung & Rensvold,

Table 4
Result of factor–ratio test for benevolence (B1... B4)

Constrained item	Referent			
	B1	B2	B3	B4
B1	1			
B2	6.457	1		
B3	8.525	.168	1	
B4	2.359	8.986	12.830	1

Non-shaded entries are significant at $p < .01$ ($df=2$).

Table 5
Result of factor–ratio test for integrity (I1... I6)

Constrained item	Referent					
	11	12	13	14	15	16
11	‡					
12	1.238	‡				
13	.987	1.748	‡			
14	4.218	7.457	7.057	‡		
15	4.607	7.153	5.796	.654	‡	
16	1.496	.203	2.695	5.614	6.345	‡

No entry is significant at $p < .01$ ($df=2$).

2002). Indeed, all other fit indices showed very minimal decrement, which is considered to support invariance irrespective of, or despite obtaining a significant chi-square statistic (Robert, Lee & Chan, 2006).

The measurement invariance literature to date is not unanimous with respect to how to proceed after establishing partial metric invariance. While several researchers advocate continuing with further tests of invariance or substantive hypotheses upon establishing partial invariance (e.g., Steenkamp & Baumgartner, 1998), others are more cautious about relying on data driven analyses that may threaten construct validity and recommend further tests only if invariance constraints are relaxed for a minority number of items on a theoretical basis and when cross-validation evidence points to their viability (e.g., Vandenberg & Lance, 2000). While the availability of data did not allow a cross-validation, given that the invariant sets consist of a sufficient number of items for both scales, we proceeded with a test of partial scalar invariance. In testing for partial scalar invariance, the intercepts of those items that are not metrically invariant across groups are left unconstrained across countries, while the intercepts of the other items are held invariant (Steenkamp & Baumgartner, 1998). Establishment of scalar invariance is a necessary condition for conducting mean comparisons across countries. Although our main focus was the cross-cultural generalizability of the operationalizations of the constructs and we had no theoretical expectation regarding mean differences, we undertook this analysis as it has been recommended as constituting a final step before proceeding into further tests of invariance which are expected to vary depending on the research question (Vandenberg & Lance, 2000).

In the first step, we constrained the intercepts of all the items for which metric invariance was established. That is, we constrained the intercepts of three items of the ability scale (A2, A3, and A4), three items of the benevolence scale (B1, B2 and B3) and all items of the integrity scale as we analyzed the scalar invariance of each scale. Chi square change statistics reported in Table 6 indicated that the intercepts associated with at least some items of these scales were non-invariant. Therefore in the second step, we tested for partial scalar invariance by constraining the intercept of a single metrically invariant item one at a time (Cheung & Rensvold, 1999). However, the sequential tests repeatedly produced statistically significant decrement in fit. Thus, we concluded that partial scalar invariance could not be established for any of the scales.

Table 6
Fit indices for the scalar invariance assessment

Scale	Model	χ^2	df	$\Delta\chi^2$	Δdf	CFI	TLI	RMSEA
Ability	Unconstrained model	179.293	31			.967	.952	.070
	Constrained model	529.771	37	350.478 *	6	.891	.867	.117
Benevolence	Unconstrained model	20.674	10			.996	.992	.033
	Constrained model	222.498	16	201.824 *	6	.916	.905	.115
Integrity	Unconstrained model	81.174	37			.984	.981	.035
	Constrained model	319.050	49	237.876 *	12	.905	.912	.075

Note: intercepts for all items of integrity, three items of ability (A2, A3, and A4) and three items of benevolence (B1, B2, and B3) scales were constrained. TLI=Tucker–Lewis Index; CFI=Comparative fit index; RMSEA=Root mean squared error of approximation.

* $p < .05$.

4. Discussion

“Leaders at all organization levels find themselves concerned with matters of multicultural relations and whether, what, and how cross-border learnings may be possible.... They grapple with questions of how far scientific social research can take us, and how the organization science ideas and methods developed in the United States and other technologically-advanced societies can be used elsewhere in the world.” (Peterson & Hunt, 1997: 203).

As stated in this quote from a review of international leadership research, leaders and scholars are seeking to determine whether constructs that have had significant explanatory power in the U.S. have similar power when doing business in other cultures. Engendering trust has been a powerful motivational driver for leaders in the U.S. (Brower et al., 2000) and this study contributes to the body of work on supervisor trustworthiness by assessing the metric equivalence of the scales developed by Mayer et al. across three samples: the US, Turkey and Singapore. Our analysis of the FOT scales supported the metric equivalence of the integrity scale, but the current operationalizations of some of the ability and benevolence items appeared to be evoking different frames of reference across the samples. Although it was possible to establish partial metric invariance with respect to these two scales (which also did not significantly decrease the internal consistency of either scale), it may be worthwhile to reconsider the proposed theoretical reasons that may have accounted for the results. Especially regarding the operationalization of the ability construct, it is our contention that items that explicitly refer to task performance are unlikely to make universal sense across cultural groups. In fact, the items that were problematic in this scale referenced the supervisor’s specific job or skills. Such operationalizations are particularly risky in the context of supervisor–subordinate relationships. We feel that the distinction between task versus contextual performance may be relevant for the cross-culturally valid operationalization of this construct.

Although we had originally expected invariance with respect to the benevolence scale, in hindsight, it appears plausible that the current items may be interpreted dissimilarly by respondents from collectivist-high power distant versus individualist-low power distant cultures. As mentioned, Turkey and Singapore are paternalistic cultures where supervisors are not only concerned about the professional but also the personal well being of their subordinates. Aycan (2001), for instance, notes that paternalistic supervisors attend weddings of employees, act as a mediator in family disputes, provide financial assistance for housing or education of children and the like. It is possible that respondents from U.S. as opposed to Turkey and Singapore used different frames of reference especially with respect to personal versus professional life domains when evaluating the benevolence of their supervisors. We propose that cross-cultural adaptations of the benevolence scale might benefit from specifying the life domain. In fact, the cross-cultural relevance or acceptability of supervisor involvement in the two life domains, however benevolent, might emerge as a culture-specific variant in the process of trust formation.

While at least partial metric invariance was established across the three cultures for three FOT scales, we found the trust scale to have suboptimal psychometric properties across the board to the extent that tests of measurement invariance were not warranted. We suggest item wording to be considered as one potential source of the problem. We recommend systematic analysis of item wording as part of future efforts at revising the Trust scale. We also propose increasing the number of items to increase the reliability of the scale. As this and the other scales are refined, we advocate the formation of a multinational team of trust and leadership scholars who work together to develop scales that are decentered (Van de Vijver & Leung, 1997). By “decentered” we mean that items are not written to reflect a single culture but are more applicable both in meaning and choice of expression to many cultures. As emphasized by Hofstede (1993), all countries have something that they call management, but its meaning differs widely from one country to another and it takes considerable insight into local history, customs, and conditions to understand it. Thus, a cross-cultural research team is needed to develop measures of FOT and trust that have conceptual equivalence across cultures.

4.1. Implications for leadership practice

Rapid globalization has pushed organizational leaders to quickly adapt management practices for a global workforce (Ryan, Chan, Ployhart, & Slade, 1999). As researchers rush to understand what management practices work in expanded operations in multiple cultures, they often overlook the need to examine the equivalence of the measures used to operationalize these constructs. As noted by Ryan et al. (1999), researchers often use an imposed-etic approach where an instrument devised in one culture is assumed to be universally applicable. In their review of cross-cultural

methodologies for organizational research, Schaffer & Riordan (2003) clearly indicate that this assumption is rarely questioned and only 17% of comparative research studies have used covariance structure analyses. It would not be far fetched to assume that the percentage of human resource practitioners or consultants who are sensitive to this problem is no higher. However, this study provides a clear example of the dangers of hasty analyses that overlook tests of equivalence. If a pseudo-etic approach were endorsed, an organizational researcher may use these same measures of FOT and trust to assess the relationship between trust in the supervisor and subordinate outcomes in countries like Turkey and Singapore. The results would not be valid and the recommendations to managers would not be meaningful (Riordan & Vandenberg, 1994). One exception perhaps is the integrity scale. The results of this study indicate that people in Singapore, Turkey and the U.S. interpret this integrity scale similarly; therefore, we can cautiously recommend that leaders who want to understand the motivational impact of a leader's integrity on significant outcomes could use this scale.

4.2. Limitations

Although the present investigation is a pioneering effort in the cross-cultural interpersonal trust literature, the contribution of our analyses should be evaluated in light of its limitations. We did not include the propensity to trust scale in our comparative analysis based on existing empirical evidence that suggested the scale needs a revision even in its original language. This was an unfortunate exclusion because, in the original model, this construct is hypothesized to be an individual difference that may be shaped by one's culture, among other things (Mayer et al., 1995). We propose that researchers move forward in two ways: develop a valid measure of propensity to trust that is not culturally bound and investigate the cultural dimensions that shape one's propensity to trust. The good news is that there has been considerable theoretical and empirical advancement in this direction in the recent years. In particular, one of the five pancultural general beliefs identified by the social axioms study carried out in 41 cultural groups by Leung et al. (Leung et al., 2002; Bond et al., 2004) is "cynicism", which refers to a negative view of human nature and a mistrust of social institutions and is proposed to have implications for altruistic and trusting behavior. We feel that this growing literature provides a promising starting point to incorporate this important variable in to the existing organizational trust and leadership paradigms.

Another limitation is with respect to the countries covered in the analysis. Due to practical concerns, we only examined three countries. Although the Turkish and the Singapore samples constituted a stringent measurement invariance test for scales developed in the U.S. context, replications with other cultural samples are necessary to support the current findings. Furthermore, we were able to access convenience samples in terms of the organizations and the participants. Although our sample sizes were moderate, future research should adopt sampling strategies that ensure representativeness of the culture to a greater extent. We also recommend that further assessments explicitly address each threat to measurement invariance as outlined by Robert et al. (2006) and make a concerted effort to control for differences in job context, organizational culture, education levels and language (translation) to isolate the effect of societal culture. While addressing all potential threats may be too much to expect from a single study, a coordinated effort by several researchers in the field would be invaluable to advance our current understanding the cultural processes.

Finally, the study was conducted with the assumption that country membership can sufficiently act as a surrogate measure for culture. Although several studies support the cultural assumptions that guided the selection of the samples (e.g., Gupta, Hanges, & Dorfman, 2002), future studies should incorporate direct measures of culture into their research design rather than assuming country membership is an assessment of cultural orientation (Schaffer & Riordan, 2003).

Appendix A

Survey items for trustworthiness and trust scales (English)

A1	My supervisor is very capable of performing his job.
A2	My supervisor is known to be successful at the things he tries to do.
A3	My supervisor has much knowledge about the work that needs to be done.
A4	I feel very confident about my supervisor's skills.
A5	My supervisor has specialized capabilities that can increase our performance.
A6	My supervisor is well qualified.

(continued on next page)

Appendix A (continued)

B1		My supervisor is very concerned about my welfare.
B2		My supervisor would not knowingly do anything to hurt me.
B3		My supervisor really looks out for what is important to me.
B4		My supervisor will go out of his way to help me.
I1	R	My supervisor's actions and behaviors are not very consistent.
I2		My supervisor has a strong sense of justice.
I3		I never have to wonder whether my supervisor will stick to his word.
I4		I like my supervisor's values.
I5		Sound principles seem to guide my supervisor's behavior.
I6		My supervisor tries hard to be fair in dealings with others.
T1	R	If I had my way, I wouldn't let my supervisor have any influence over issues that are important to me.
T2		I would be willing to let my supervisor have complete control over my future in the organization.
T3	R	I really wish I had a good way to keep an eye on my supervisor.
T4		I would be comfortable giving my supervisor a task or problem which was critical to me, even if I could not monitor their actions.

Note: (R) denotes reverse-coded item.

References

- Adler, N. J., & Jelinek, M. (1986). Is "organizational culture" culture bound? *Human Resource Management*, 25, 73–90.
- Anderson, S. E., & Williams, L. J. (1996). Interpersonal, job and individual factors related to helping processes at work. *Journal of Applied Psychology*, 81, 282–296.
- Aycan, Z. (2001). Human resource management in Turkey: Current issues and future challenges. *International Journal of Manpower*, 22(3), 252–261.
- Aycan, Z., Kanungo, R. N., Mendonca, M., Yu, K., Deller, J., Stahl, G., et al. (2000). Impact of culture on human resource management practices: A ten country comparison. *Applied Psychology: An International Review*, 49, 192–220.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246.
- Brower, H. H., Schoorman, F. D., & Tan, H. H. (2000). A model of relational leadership: The integration of trust and leader–member exchange. *Leadership Quarterly*, 11, 227–250.
- Bond, M. H., Leung, K., Au, A., de Carrasquel, S. R., Murakami, F., Yamaguchi, S., et al. (2004). Culture-level dimensions of social axioms and their correlates across 41 cultures. *Journal of Cross-Cultural Psychology*, 35, 548–570.
- Casimir, G., Waldman, D. A., Bartram, T., & Yang, S. (2006). Trust and the relationship between leadership and follower performance: Opening the black box in Australia and China. *Journal of Leadership and Organizational Studies*, 12(3), 68–84.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255.
- Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, 25, 1–27.
- Currall, S. C., & Judge, T. A. (1995). Measuring trust between organizational boundary role persons. *Organizational Behavior and Human Decision Processes*, 64, 151–170.
- Davis, J. H., Schoorman, F. D., Mayer, R., & Tan, H. H. (2000). Trusted unit manager and business unit performance: Empirical evidence of a competitive advantage. *Strategic Management Journal*, 21, 563–576.
- Dirks, K. T. (1999). The effects of interpersonal trust on work group performance. *Journal of Applied Psychology*, 84, 445–455.
- Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic finding and implications for research and practice. *Journal of Applied Psychology*, 87, 611–628.
- Doney, P. M., Cannon, J. P., & Mullen, M. (1998). Understanding the influence of national culture on the development of trust. *Academy of Management Review*, 23, 601–620.
- Graen, G. B., & Uhl-Bien, M. (1995). Relationship-based approach to leadership: Development of leader–member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *Leadership Quarterly*, 6, 219–247.
- Green, S. G., Anderson, S. E., & Shivers, S. L. (1996). Demographic and organizational influences on leader–member exchange and related work attitudes. *Organizational Behavior and Human Decision Processes*, 66, 203–214.
- Gupta, V., Hanges, P. J., & Dorfman, P. (2002). Cultural clusters: Methodology and findings. *Journal of World Business*, 37, 11–15.
- Hofstede, G. (1993). Cultural constraints in management theories. *Academy of Management Executive*, 7, 81–94.
- Hofstede, G. (1980). *Culture's consequences*. Beverly Hills: Sage Publications.
- Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and philosophical ethics. *Academy of Management Review*, 20, 379–403.
- House, R. J. (1976). Theory of charismatic leadership. In J. G. Hunt & L. L. Larson (Eds.), *In Leadership: The cutting edge* Carbondale, Illinois: University Press.
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage Publications.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.
- Kabasakal, H., & Bodur, M. (2002). Arabic cluster: A bridge between East and West. *Journal of World Business*, 37, 40–54.

- Kramer, R. M., & Tyler, T. R. (Eds.). (1996). *Trust in organizations*. Newbury Park, CA: Sage Publications.
- Liden, R. C., & Maslyn, J. M. (1998). Multidimensionality of leader–member exchange: An empirical assessment through scale development. *Journal of Management*, 24, 43–50.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32, 53–76.
- Leung, K., Bond, M. H., de Carrasquel, S. R., Muñoz, C., Hernández, M., Murakami, F., et al. (2002). Social axioms: The search of universal dimensions of general beliefs about how the world functions. *Journal of Cross-Cultural Psychology*, 33, 286–302.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224–253.
- Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, 84, 123–136.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709–734.
- Mayer, R. C., & Gavin, M. (2005). Trust for management and performance: Who minds the shop while the employees watch the boss? *Academy of Management Journal*, 48(5), 874–888.
- Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, 128, 3–72.
- Peterson, M. F., & Hunt, J. G. (1997). International perspectives on international leadership. *Leadership Quarterly*, 8, 203–231.
- Pillai, R., Schriesheim, C. A., & Williams, E. S. (1999). Fairness perceptions and trust as mediators for transformational and transactional leadership: A two-sample study. *Journal of Management*, 25, 897–936.
- Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *Leadership Quarterly*, 1, 142–207.
- Putti, J. M., & Chong, T. (1987). American and Japanese management practices in their Singapore subsidiaries. *Asia Pacific Journal of Management*, 2, 103–114.
- Riordan, C. M., & Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management*, 20, 643–671.
- Robert, C., Lee, W. C., & Chan, K. Y. (2006). An empirical analysis of measurement equivalence with the INDCOL measure of individualism and collectivism: Implications for valid cross-cultural inference. *Personnel Psychology*, 59, 65–99.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23, 393–404.
- Ryan, A. M., Chan, D., Ployhart, R. E., & Slade, L. A. (1999). Employee attitude surveys in a multinational organization: Considering language and culture in assessing measurement equivalence. *Personnel Psychology*, 52, 37–58.
- Scandura, T. A., & Schriesheim, C. A. (1994). Leader–member exchange and supervisor career mentoring as complementary constructs in leadership research. *Academy of Management Journal*, 37, 1588–1602.
- Schaffer, B. S., & Riordan, C. M. (2003). A review of cross-cultural methodologies for organizational research: A best-practices approach. *Organizational Research Methods*, 6, 169–215.
- Schoorman, F. D., Mayer, R. C., & Davis, J. H. (1996). Empowerment in veterinary clinics: The role of trust in delegation. *Paper presented at the 11th Annual Meeting for Industrial and Organizational Psychology, San Diego, CA, April*.
- Schriesheim, C. A., Neider, L. L., & Scandura, T. A. (1998). Delegation and leader–member exchange: Main effects, moderators, and measurement issues. *Academy of Management Journal*, 41, 298–318.
- Seers, A. (1989). Team-member exchange quality: A new construct for role-making research. *Organizational Behavior and Human Decision Processes*, 43, 118–135.
- Smith, P. S., Fischer, R., & Sale, N. (2001). Cross-cultural organisational psychology. In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology, Vol. 16* (pp. 147–193). New York: John Wiley.
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78–90.
- Tan, H. H., & Chee, D. (2005). Understanding interpersonal trust in a Confucian influenced society: An exploratory study. *International Journal of Cross Cultural Management*, 5, 197–212.
- Tan, H. H., & Wee, G. (2002). The role of rhetoric content in charismatic leadership: A content analysis of a Singaporean leader's speeches. *International Journal of Organization Theory and Behavior*, 5, 317–342.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, Co: Westview Press.
- Triandis, H. C., & Gelfand, M. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, 74, 118–128.
- Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks: Sage.
- 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, 4–69.
- Whitener, E. M., Brodt, S. E., Korsgaard, M. A., & Werner, J. M. (1998). Managers as initiators of trust: An exchange relationship framework for understanding managerial trustworthy behavior. *Academy of Management Review*, 23, 513–530.