A Multimethod Multitrait Validity Assessment of Self-Construal in Japan, Korea, and the United States

MARY J. BRESNAHAN
TIMOTHY R. LEVINE
SACHIYO MORINAGA SHEARMAN
SUN YOUNG LEE
Michigan State University
CHEONG-YI PARK
Inha University
TORU KIYOMIYA
Seinan Gakuin University

A large number of previous studies have used self-construal to predict communication out-comes. Recent evidence, however, suggests that validity problems may exist in self-construal measurement. The current study conducted a multimethod multitrait (Campbell & Fiske, 1959) validation study of self-construal measures with data (total N = 578) collected in Korea (N = 200), Japan (N = 212), and the U.S. (N = 166). The data showed that the Singelis (1994) Self-Construal Scale, the Cross, Bacon, and Morris (2000) Relational Interdependent Self-Construal Scale (RISC), and the Kuhn and McPartland (1954) Twenty Statements Test (TST) lacked convergent and discriminant validity, both pan-culturally and within each of the three countries included in the study. Scores on the TST were not significantly related to scores on the self-construal scales, and the various self-construal measures correlated more highly with measures of communication directness than with alternative measures of the same type of self-construal. Substantial method effects were also observed. The results were tested for both 2- and 3-dimensional models of self-construal and for refined scales and scales with all items retained. The results of all analyses were inconsistent with the claim that self-construal measures are construct valid.

Self-construal (Markus & Kitayama, 1991) involves linking different aspects of self-concept with specific cultural differences, especially individualism and collectivism. At least three types of self-construals have been identified. Independent self-construal is based on personal autonomy and uniqueness from others, collective interdependent self-construal describes one's position in the group and the maintenance of group harmony, and relational interdependent self-construal is defined by connection to others in close committed relationships (Cross, Bacon, & Morris, 2000).[1] Self-construal is often viewed as individual-level individualism–collectivism (e.g., Kim, 2002; Oyserman, Coon, & Kemmelmeier, 2002), a mediator between national-level culture and outcome variables

(e.g., Gudykunst et. al., 1996; Kim, Hunter, Miyahara, Horvath, & Bresnahan, 1996), and as mechanism for explaining within-culture variance (e.g., Gudykunst & Lee, 2003; Kim & Raja, 2003). Self-construal has been measured or induced in a variety of ways including Kuhn and McPartland's (1954) Twenty Statements Test (TST)—which involves the coding of open-ended self-descriptions, pronoun and other priming tasks (e.g., Gardner, Gabriel, & Lee, 1999)—and several Likert-type self-report self-construal scales (e.g., Cross, Bacon, & Morris, 2000; Gudykunst et al., 1996; Leung & Kim, 1997; Singelis, 1994).

The self-construal constructs have been enthusiastically received as an alternative to country-level cultural comparison, and research on the topic has grown exponentially during recent years. Self-construals are touted as a major theoretical advance (Kim, 2002). They have been shown to be associated in predictable ways with an impressive number of outcome measures (Gudykunst & Lee, 2003), and interest in self-construals spans a number of social scientific disciplines including communication, sociology, psychology, management, and education (Levine et al., 2003).

Until recently, few studies had systematically evaluated the validity of self-construal measures, and early validation studies concluded that self-construal scales were valid (e.g., Hackman, Ellis, Johnson, & Staley, 1999; Singelis, 1994). More recently, advocates of self-construal have argued for the more moderate position that the validity of self-construal is a "work in progress" and validity will eventually be established through repeated research (Kim & Raja, 2003). Nevertheless, a growing number of studies report results inconsistent with the premises underlying the self-construal construct (e.g., Bresnahan, Levine, & Chiu, 2004; Levine et al., 2002; Park, 2001; Park & Levine, 1999; Rao, Singhal, Ren, & Zhang, 2001; Sato & Cameron, 1999). Thus, whereas the self-construal constructs complement the dominant theoretical view of individualism–collectivism, a growing number of anomalous findings justify greater attention to validity concerns.

Recently, four articles offered compelling empirical evidence of validity problems. Oyserman et al. (2002) reported an extensive meta-analysis of national differences in individualism and collectivism that included self-construal measures. Oyserman et al. reported that predicted differences were often not evident for most county comparisons, the literature was characterized by substantial heterogeneity of effects, and that scale item content was a strong moderator of country differences. Based on these results, Osyerman et al. cautioned against the uncritical use of self-construal, and described the state of the literature as "tentative" (p. 87). [2] Similarly, Levine et al. (2003a, 2003b), using a variety of methodologies including meta-analysis, demonstrated that average effect sizes in self-construal research were often very small and widely variant across studies—and that many findings were in the wrong direction from

hypotheses. Evidence from factor analyses further indicated that items comprising the various self-construal scales did not produce stable factors. Finally, Grace and Kramer (2003) and Levine et al. (2003a) presented data suggesting that various measures of self-construal failed to converge. Grace and Kramer reported that Twenty Statements Test (TST) scores and self-construal scores did not correlate substantially, and Levine et al. found the self-construal scales were unaffected by independent and interdependent priming inductions.

There is also conceptual and empirical debate about the number and nature of self-construal dimensions, as well as a growing dissatisfaction with the two-factor model originally described by Markus and Kitayama (1991) and subsequently by Singelis (1994) and Kim (2002). Several studies suggest that self-construal should be thought of as having more than two dimensions, but there does not appear to be consensus about the number or nature of additional dimensions (e.g., Cross et al., 2000; Fiske, 2002; Grace & Cramer, 2003; Kashima, 2002; Kashima & Hardie, 2000; Levine et al., 2003). Cross et al. made a compelling case for adding a relational interdependent dimension, but it is unclear if three dimensions are sufficient (cf. Fiske, 2002).

Together, these disconfirming and contradictory results either reflect problems in the way that self-construal has been measured, deeper conceptual problems, or both (Levine et al., 2003a, 2003b). Arguing on the basis of the Cronbach and Meehl (1955) nomological network approach to construct validation, Gudykunst and Lee (2003) contended that since self-construal scales have been shown to correlate with outcome measures as predicted, the scales must be construct valid. Levine et al. (2003a, 2003b), however, speculated that such correlations might be spurious and attributable to item content confounds. Gudykunst and Lee (2003) and Levine et al. (2003a, 2003b) agreed that "multi-trait, multi-method procedures would be ideal ways to establish validity, and future attempts to establish the validity of self construals should use these procedures" (Gudykunst & Lee, 2003, p. 264). The Campbell and Fiske (1959) multimethod, multitrait matrix (MMMT) approach is therefore used in this study to provide a rigorous test of construct validity of self-construal scales.

MULTIMETHOD MULTITRAIT CONSTRUCT VALIDATION

Campbell and Fiske (1959) observed that the results in many literatures are limited to a single operational definition for a given construct. In such cases, systematic variance can be affected by either "measurement features, trait content or by both" (Campbell & Fiske, 1959, p. 91), because the constructs being measured are confounded with the methods

used in measurement. As an alternative, they advocated "multiple operationalism and use of method triangulation" to establish construct validity (Campbell & Fiske, 1959, p. 101). A solution to construct–method confounds involves crossing a minimum of two independent measurement procedures with at least two different constructs to test construct validity. Specifically, Campbell and Fiske's (1959) MMMT matrix crosses constructs with methods to generate four kinds of correlations. These include same trait–same method, same trait–different method, different trait–same method, and different trait–different method correlations. These correlations provide evidence for reliability, convergent validity, predictive validity, and discriminant validity.

Conceptually, a reliability coefficient is a correlation between a variable and itself. These are listed along the primary diagonal of the MMMT matrix and can be thought of as same trait–same method correlations. Same trait–different method correlations reflect the association between different measures of the same construct. Strong and positive correlations among different measures of the same construct provide evidence for convergent validity. The principle is that alternative measures of the same construct should converge. Campbell and Fiske suggested that validity was indicated by a correlation of .50 or higher. Lower correlations, even if they are statistically significant, suggest that there may be problems for construct validity. Different trait–different method correlations reflect associations between different constructs that are not confounded with a common method. These associations may be positive, negative or trivial depending on theory-based hypotheses; to the extent that the associations are theory consistent, they provide evidence for predictive validity. Finally, different trait–same method correlations are associations between different constructs that are measured with a common method. To the extent that the different trait–same method correlations are smaller than the same trait–different method correlations, evidence for convergent and discriminant validity exists. When the correlations among different traits using the same method are larger than for the same traits using different methods, evidence for method effects exists. Thus, the MMMT approach offers rigorous tests of convergent, predictive, and discriminant validity by removing construct-method confounds.

In a MMMT matrix, the constructs that are selected should be based on "theoretically predicted associations" (Campbell & Fiske, 1959, p. 100), but the different constructs must also be conceptually distinct. In the current study, direct and indirect communication styles were chosen as additional constructs, because directness of communication has been both theoretically and empirically linked with self-construals in previous research (e.g., Singelis, 1994, Gudykunst et al., 1996) and aspects of self-concept and communication style are clearly different constructs (Kim &

Raja, 2003; Levine et al., 2003b). More specifically, the independent constructs measured in the current study include independent self-construal, collective interdependent self-construal, relational interdependent self-construal, direct communication style, and indirect communication style. These constructs are each measured using two different methods: close-ended Likert scales and coded open-ended responses. The Likert-type scales include the Singelis (1994) Self-Construal Scale, Cross et al.'s (2000) Relational Interdependent Self-Construal Scale (RISC), and direct and indirect communication style scales developed by the authors. The open-ended, free listing responses are assessed using the Kuhn and McPartland (1953) Twenty Statements Test and also by coding participants' responses to scenarios for directness.

RESEARCH PREDICTIONS

Hypotheses for Cultural Differences

Whereas the conceptualization of self-construal maintains that every person possesses both independent and interdependent selves (Gudykunst et al., 1996, Kim, 2002), many self-construal theorists contend that one type of self-construal should predominate in most situations for members of specific cultural groups (e.g., Bresnahan, Levine, & Chiu, 2004; Gudykunst & Lee, 2003; Haberstroh, Oyserman, Schwartz, Kuhnen, & Ji, 2002; Markus & Kitayama, 1991; Oyserman et al., 2002). [3] To the extent that self-construal scores and communication styles reflect anticipated cross-cultural differences, the following hypotheses are warranted from the previous literature (e.g., Gudykunst et al., 1996, Markus & Kitayama, 1991, Singelis, 1994):

H1: Scores on independent self-construal (H1a) and direct communication style scales (H1b) will be higher for American participants compared with Japanese and Koreans.

H2: Scores on collective interdependent self-construal (H2a) and indirect communication style scales (H2b) will be higher for Korean and Japanese participants compared with Americans.

H3: Scores on the TST will be higher in independence (H3a) and lower in collective interdependence (H3b) for American participants compared with Japanese and Koreans.

H4: Responses to the open-ended scenarios will show that Japanese and Korean participants are less direct in communication compared with Americans.

Cross et al. (2000) differentiate relational and collective-based interdependence in this way:

In collectivism-based interdependence, the individual's position in the group or situation dictates behavior; therefore, knowing one's place, behaving according to one's role, and putting the needs of the group before one's own needs are central dictums that shape the self-construal. *This group-oriented notion of interdependence, however, does not adequately describe the relationship-centered conception of interdependence that characterizes North Americans* (p. 792, emphasis added).

Thus, Cross et al. imply that collective-based interdependence reflects a type of self-construal associated with cultural-level collectivism whereas a relationship-centered type of interdependent self-construal is common in the U.S. Consistent with this, Kashima et al. (1995) compared participants from five countries (Australia, mainland U.S., Hawaii, Japan, Korea) and found that Japanese exhibited the lowest scores on relatedness while American women exhibited the highest score. Kanagawa, Cross, and Markus (2001) also found that Americans described themselves significantly more relationally interdependent than Japanese, and Bresnahan et al. (2004) found that participants from the U.S. exhibited higher relational scores than Japanese and Koreans. Given these arguments and findings, the following hypothesis is warranted:

> H5: Scores for relational interdependent self-construal will be higher for participants from the U.S. compared with Japanese and Koreans.

Hypotheses for Convergent Validity

Convergent validity is based on the idea that there should be strong and positive correlations between different measures of the same construct. Previous studies by Triandis and colleagues provided evidence of convergence between self-identity measured with Likert scales and concepts of self elicited through the TST. For example, Triandis, Chan, Bhawuk, Iwao and Sinha (1995) argued that self-construal can "be measured with content analysis of the responses to twenty statements or scales that measure independence versus interdependence developed by Singelis (1994) and Gudykunst et al. (1994)" (pp. 462–463). Watkins, Yau, Dahlin and Wondimu (1997) similarly compared TST with independent measures of self-concept. Sunar (1999), over the course of four studies, compared TST with a scaled measure of individualism and collectivism. [4] Further, given that the TST involves coding self-descriptions, it has very high face validity. Common in the literature, therefore, is the view that the TST and the self-construal scales provide alternative measures of the same constructs:

> H6: Scores on the independent self-construal scale will correlate strongly and positively (i.e., $r > .50$) with independent answers on the TST.

H7: Scores on the interdependent self-construal scale will correlate strongly and positively with interdependent answers on the TST.

H8: Scores on RISC will correlate strongly and positively with relational answers on the TST.

Similarly, although not the primary focus here, the measures of communication directness used in the current study should exhibit convergent validity. Therefore, we proposed additional hypotheses:

H9: Scores on the direct communication scale will correlate strongly and positively with coded directness scores for the scenarios.

H10: Scores on the indirect communication scale will correlate negatively with directness scores on the scenarios.[5]

Hypotheses for Predictive Validity

Different trait–different method correlations reflect associations between different constructs that are not confounded with a common method. Different trait–same method correlations also show the relationship among different constructs, but are confounded by a common method. These associations may be positive, negative, or trivial depending on the similarity of the constructs, but they should be substantially less in magnitude than the same trait–different method correlations.

The relationship between construal of the self and how one communicates with others has been extensively discussed in the self-construal literature. Markus and Kitayama (1991) suggested that independent self-construal was associated with direct communication style while interdependent self-construal was associated with indirect communication style. Kim, Sharkey, and Singelis (1994) similarly described that people with strong independent self-construal emphasized a preference for clear and explicit communication, while people high in interdependent self-construal showed a preference for verbal indirectness. Gudykunst et al. (1996) reported that independent self-construal predicted precision of communication while interdependent self-construal predicted indirect communication and greater interpersonal sensitivity. In addition, a number of studies have provided evidence that Japanese prefer indirect, implicit, and more ambiguous communication style compared to Americans (e.g., Hasegawa, 1996; Nomura and Barnlund, 1983; Thompson, Klopf, & Ishii, 1991). Other studies have shown that Americans were more direct than Koreans (Kim & Bresnahan, 1996; Kim et al., 1996). Thus, there is ample theoretical reasoning and empirical evidence connecting self-construal with direct and indirect communication style. Based on the MMMT model and the self-construal literature, it is possible to make the following claims as tests of predictive validity:

H11: Scores on independent self-construal will correlate moderately and positively with scaled direct communication (H11a) and coded direct communication (H11b) and negatively with the scaled indirect communication (H11c).

H12: Collective interdependent self-construal will correlate negatively with scaled direct communication (H12a) and coded direct communication (H12b) and positively with scaled indirect communication (H12c).

H13: Independent scores on the TST will correlate moderately and positively with scaled direct communication (H13a) and coded direct communication (H13b) and negatively with scaled indirect communication (H13c).

H14: Collective interdependent scores on the TST will correlate moderately and negatively with scaled direct communication (H14a) and coded direct communication (H14b) and positively with scaled indirect communication (H14c).

H15: Relational interdependent scores on the TST will correlate moderately and negatively with scaled direct communication (H15a) and coded direct communication (H15b) and positively with scaled indirect communication (H15c).

Hypotheses for Discriminant Validity

To the extent that different trait–same method correlations are smaller than the same trait–different method correlations, evidence for convergent and discriminant validity is found. Thus, based on the MMMT model applied to self-construals, it is possible to make the following hypotheses:

H16: Scores from different measures of the same construct will correlate more highly with each other than with other constructs.

More specifically, this suggests the following expected correlations between constructs in the current study:

H16a: Scores on the independent self-construal scale will correlate more highly with independent responses on the TST than with measures of other dimensions of self-construal or with measures of communication style.

H16b: Scores on the collective interdependent self-construal scale will correlate more highly with interdependent responses on the TST than with measures of other dimensions of self-construal or with measures of communication style.

H16c: Scores on the relational interdependent self-construal scale will correlate more highly with relational responses on the TST than with measures of other dimensions of self-construal or with measures of communication style.

H16d: Scores on direct communication style scale will correlate more highly with coded directness than with measures of self-construal.

H16e: Scores on the indirect communication style scale will correlate more nega-
tively with coded directness than with measures of self-construal.

Measurement Variance

Statistically significant associations can also result merely from the use of common measurement techniques (Nunnally & Bernstein, 1994). Levine et al. (2003a, 2003b) speculate that much of the evidence for self-construal scales correlating with outside measures might be attributable to same-method artifacts. Thus, it is important to test for the presence of measurement variance. Measurement variance is indicated by stronger correlations between similar measures of different constructs compared to different measures of the same constructs.

RQ1: To what extent are correlations between different constructs affected by common measurement methods?

METHOD

Participants and Procedures

Five hundred and seventy-eight participants were included in this study with 166 from the U.S. (55 males, 111 females), 212 from Japan (109 males, 97 females, 6 unclassified), and 200 from Korea (92 males, 102 females, 6 unclassified). American participants were enrolled in an introductory communication class at Michigan State University. They ranged in age from 18 to 29 years with a mean age of 20.01. They participated in the study as part of a course research requirement. Japanese participants were voluntarily recruited from introductory communication classes at Seinan Gakuin University in Tokyo. They ranged in age from 17 to 25 years with an average age of 19.85. Korean participants, ranging in age from 20 to 28 years with an average age of 23.60, were voluntarily recruited from communication classes at Inha University in Incheon City. The research collaborators in Korea and Japan had received their training at universities in the U.S. and were aware of the importance of protecting human subjects and obtaining informed consent for research. The study was institutional review board approved in the U.S.

Participants answered all open-ended items before completing the five scales to minimize priming. The first open-ended measure was the TST. Following this, participants read four scenarios and were asked whether they would respond to the problem posed in the scenario—and, if so, what were the exact words that they would say to a stranger who had

offended them in some way. The scenarios were developed by the authors and piloted for their realism, plausibility, and cultural appropriateness.[6] Participants were asked to respond to all four scenarios to enhance reliability. The scenarios are included in Appendix A. The five scales followed the open-ended responses including the 12 item Singelis (1994) independent self-construal scale, the 12 item Singelis interdependent scale,[7] the 11 item Cross et al. (2000) relational interdependent scale, a 13 item scale to measure direct communication developed by the authors in consultation with the literature on communication style (cf. Gudykunst et al., 1996; Nelson, Batal, & El Bakar, 2002), and a 12 item scale to measure indirect communication style also developed by the authors. All scales had 5-point Likert format with 1 = *strongly disagree* and 5 = *strongly agree*. The direct and indirect communication scales are included in Appendix B. The questionnaire was developed in English and then translated into Korean and Japanese. A back-translation procedure by paid translators who were independent of the project was used to ensure the quality of the translation.

Coding

Six trained coders (two from each of the three language groups) worked independently coding answers on the open-ended TST and the four scenarios. All coding was conducted in the language of the respondents by bilingual coders. Responses on the TST were coded for autonomy and unique characteristics (independent self-construal), group connection (collective interdependent self-construal), and references to committed, close relationships (relational interdependent self-construal).[8] Intercoder reliability for the TST was Cohen's κ = .90 for the American coders based on 53 surveys coded by both, Cohen's κ = .92 for the Japanese coders based on 52 surveys coded by both, and Cohen's κ = .93 for the pair of Korean coders based on 55 surveys coded in common. Coders also read the open-ended responses for each of the four scenarios and then made a global evaluation of each message ranking each response using the categories where 1 = *choose not to respond*, 2 = *an indirect message such as hinting about something*, 3 = *a direct but mitigated message*, 4 = *a direct message*, 5 = *a hostile direct message*.[9] A higher mean score indicated more direct communication. Intercoder reliability for 50 scenarios coded in common by each pair of coders was Cohen's κ = .90 for the US coders, Cohen's κ = .93 for the Japanese coders, and Cohen's κ = .95 for the Korean coders.

Scale Evaluation

Item analysis for each scale was conducted using corrected item-total correlations. All items were expected to have a corrected correlation of .40 or greater both pan-culturally and within each country, as well as

contribute positively to scale reliability as the standards for inclusion of an item in the final scale. Based on these criteria, eight items were found to be reliable both pan-culturally and within each country for the direct communication scale, pan-cultural α = .82, U.S. = .83, Japan = .70, and Korea = .74. Seven items were found to measure indirectness, pan-cultural α for indirectness = .77, U.S. = .80, Japan = .72, Korea = .83.[10] Six items were found to measure relational interdependent self-construal (RISC) with a pan-cultural α of .74, U.S. = .78, Japan = .68, and Korea = .74.

The Singelis scale posed a problem for obtaining unidimensionality. Different items correlated differently in each of the three countries. Two factors as theoretically predicted by Singelis (1994) could not be obtained using confirmatory factor analysis (Hunter & Gerbing, 1982). In order to obtain a set of items that would show acceptable reliability in the three countries, the interitem correlation cutoff had to be lowered to .25 or greater. This standard was less than ideal and suggests that there may be problems with scale validity (cf. Levine et al., 2003a). Nine of the original twelve independent self-construal items were found to be reliable both pan-culturally and within each country, overall pan-cultural α = .60, in the US = .58, Japan = .58, and Korea = .54. Six items for collective interdependent self-construal were reliable both pan-culturally and within each country, pan-cultural α = .50, US = .51, Japan = .50 and Korea = .53. These reliabilities, while lower than desirable, were the best that could be obtained in the current study for the Singelis self-construal scale. An examination of the literature suggests that low reliabilities are not atypical for the Singelis scale (e.g., Grace & Cramer, 2003; Levine et al., 2003; Vohs & Heatherton, 2002).

RESULTS

A series of one-way ANOVAs showed significant between-culture differences on all measures except for the TST. F-tests and the means and standard deviations by country for all variables are presented in Table 1.[11] Planned t-tests (two-tailed) were used to determine which country results were significantly different in order to test H1–5. For each t test below, the effect size is reported as r.[12] The data were consistent with the first hypothesis. Participants in the U.S. had higher independent self-construal than either Koreans or Japanese (effect size of r = .30 and r = .10 respectively). Americans were more direct in their scaled communication style than Koreans and Japanese, r = .34 and r = .21 respectively. The data were partially consistent with the second hypothesis. Americans had lower collective interdependent self-construal compared to Koreans, r = .41, while Americans and Japanese did not differ significantly (r = .06). Consistent with H2b, both Koreans and Japanese exhibited significantly higher

TABLE 1
Mean Scores by Country for the Revised Scales

	U.S. (n = 166)		Japan (n = 212)		Korea (n = 200)	
	M	SD	M	SD	M	SD
Independent self-construal $F(1, 565) = 18.469, p < .001, eta^2 = .06$	3.65_a	0.49	3.51_b	0.52	3.35_c	0.48
Interdependent self-construal $F(1, 565) = 151.931, p < .001, eta^2 = .15$	3.46_a	0.46	3.40_a	0.57	3.91_b	0.53
Relational self-construal $F(1, 565) = 27.155, p < .001, eta^2 = .09$	3.87_a	0.57	3.38_b	0.62	3.71_c	0.55
TST independent $F(1, 537) = 1.01, p = .37,$ N.S.	17.13	2.24	17.07	1.99	17.30	2.06
TST interdependent $F(1, 535) = 3.338, p < .05, eta^2 = .01$	2.66_a	2.14	2.86_a	1.91	2.41_b	1.96
TST relational $F(1, 537) = 12.883, p < .001, eta^2 = .05$	0.20_a	0.44	0.05_b	0.33	0.28_a	0.54
Direct communication style $F(1, 565) = 34.299, p < .001, eta^2 = .11$	3.97_a	0.46	3.52_b	0.55	3.60_b	0.55
Indirect communication style $F(1, 565) = 57.988, p < .001, eta^2 = .17$	2.26_a	0.51	2.81_b	0.53	2.92_b	0.61
Scenario $F(1, 565) = 25.708, p < .001, eta^2 = .08$	2.56_a	0.83	2.11_b	0.74	2.54_a	0.74

NOTE: Different subscripts in a row indicate significant difference at $p < .05$.
Means for independent, interdependent, relational self-construal, direct communication style, and indirect communication style are based on a 5-point Likert scale with larger values indicating higher scores on the construct.

indirect communication style compared to Americans, $r = .46$ and $r = .50$ respectively. In contrast to what was expected in H3a, no differences were found for independent responses on the TST.[13] The data were not consistent with H3b. Japanese exhibited significantly higher interdependent TST compared to Koreans, $r = .12$, but were not different from Americans, $r = -.05$. As predicted in H4, Japanese opted for less direct responses to the coded scenarios compared to Koreans and Americans, $r = .28$ and $r = .28$ respectively; however, there was no difference between Koreans and Americans, $r = .01$. Thus, the data were only partially consistent with H4. Finally, the data were consistent with H5. Americans had significantly

TABLE 2
Pan-cultural Correlations and Reliabilities for Revised Scales

	INDSC	INTSC	RISC	TSTIND	TSTINT	TSTREL	DCS	ICS	SCEN
Independent SC	(.67)								
Interdependent SC	.08	(.56)							
Relational SC	.13*	.34**	(.74)						
TST independent	.05	.04	-.11*	(.93)					
TST interdependent	-.03	-.07	.09*	-.98**	(.91)				
TST relational	-.11*	.07	.11*	-.23**	.03	(.90)			
Direct com. style	.48**	.09*	.21**	.02	-.02	.03	(.85)		
Indirect com.style	-.30**	.22**	-.05	.02	-.02	-.01	-.50**	(.83)	
Scenario	.22**	.07	.11*	.08	-.09*	.07	.19**	-.09*	(.93)

NOTE: *$p < .05$, ** $p < .001$

higher RISC scores than either Koreans or Japanese, $r = .14$ and $r = .37$ respectively. Koreans were significantly higher in RISC compared to Japanese, $r = .27$. Both Koreans and Americans indicated stronger relational identities on TST compared to Japanese, $r = .25$ and $r = .19$ respectively.

Results for Convergent Validity

Convergent validity coefficients (i.e., correlations between alternate measures of the same construct) were expected to be substantial if self-construal measures were valid. In order to be valid, independent self-construal should correlate significantly and robustly with independent responses on the TST while collective interdependent self-construal should correlate with interdependent responses on the TST and relational interdependent self-construal with relational answers on the TST. As shown in Table 2, the data were not consistent with any of these hypotheses. Within each country, (as shown in Tables 3, 4, and 5) independent self-construal also failed to correlate with TST independent responses as predicted in hypothesis 6. Collective interdependent self-construal failed to correlate with interdependent responses on TST as predicted in H7. Regarding H8, a significant but trivial correlation was observed for relational interdependent self-construal and responses on TST. (The correlation of $r = .11$ did not approach the $r = .50$ criterion for convergent validity.) These results show that there is no evidence for convergent validity between two measures (scales and TST) that have been widely used to measure self-construal in previous research.

TABLE 3
Correlations and Reliabilities for Revised Scales in the U.S.

	INDSC	INTSC	RISC	TSTIND	TSTINT	TSTREL	DCS	ICS	SCEN
Independent SC	(.58)								
Interdependent SC	.16*	(.51)							
Relational SC	.04	.24**	(.78)						
TST independent	.11	.01	-.20*	(.90)					
TST interdependent	-.06	-.02	.19**	-.98**	(.90)				
TST relational	-.25**	.06	.06	-.30**	.10	(.90)			
Direct comm. style	.53**	.14	.10	.14	-.13	-.10	(.78)		
Indirect com. style	-.26**	.08	.03	.03	-.02	.02	-.61**	(.80)	
Scenario	.27**	.07	-.06	.17*	-.15	-.11	.19**	-.09	(.90)

NOTE: *$p < .05$, ** $p < .001$

The second part of convergent validity is based on the specification (H9) that direct communication style will correlate positively with responses obtained for the open-ended scenarios while indirect communication style will correlate negatively (as in H0). Whereas significant correlations were found between communication style and scenario responses, neither of these correlations approached the standard for validity suggested by Campbell and Fiske (correlations are $r = .19$, $p < .001$ for direct communication with scenario and $r = -.09$, $p < .05$ for indirect communication with scenario). Similar results were found both pan-culturally and within each country. The upper bounds of the 95% confidence intervals surrounding the mono-construct, hetero-measure correlations were in every case substantially below the criteria of $r > .50$. Therefore, the data were inconsistent with H6–10.

Results for Predictive Validity

The pan-cultural results relevant to H11–15 are presented in Table 2, and the culture specific results are presented in Tables 3–5. As can be seen in these tables, the data were consistent with H11. As predicted, independent self-construal correlated positively with scaled direct communication style, r (577) = .48, $p < .001$, and coded directness, r (577) = .22, $p < .001$, while it correlated negatively with indirect communication style, r (577) = -.30, $p < .001$. These findings were generally consistent both across and within the three cultures. The data were only partially consistent with

TABLE 4
Correlations and Reliabilities for Revised Scales in Japan

	INDSC	INTSC	RISC	TSTIND	TSTINT	TSTREL	DCS	ICS	SCEN
Independent SC	(.58)								
Interdependent SC	.13	(.50)							
Relational SC	.23**	.36**	(.68)						
TST independent	-.03	.04	-.19**	(.92)					
TST interdependent	.02	-.05	.20**	-.99**	(.92)				
TST relational	.01	.02	.05	-.19**	.05	(.92)			
Direct com. style	.42**	.07	.19**	-.03	.03	.07	(.68)		
Indirect com. style	-.27**	.15*	-.02	.05	-.03	.01	-.29**	(.72)	
Scenario	.23**	-.02	.06	.05	-.06	. 07	.19**	-.01	(.93)

NOTE: *$p < .05$, ** $p < .001$

H12. As predicted, collective interdependent self-construal correlated positively with scaled indirect communication style, r (577) = .22, $p < .001$, but correlated near zero with coded directness, r (577) = .07. These findings were somewhat variable within-country. The data for the TST were not consistent with any part of H13–15. All correlations between TSTs with direct communication style, indirect communication style, and coded directness were near zero. These data show that all measures of self-construal, with the possible exception the independent self-construal scale, fail to meet minimal standards for predictive validity. As discussed later, it is plausible that the predictive power for the independent self-construal scale may be attributed to item content confounds.

Results for Discriminant Validity

The data were not consistent with any part of H16. Scores from the alternative measures of self-construals (scales and TST) failed to correlate more highly with each other than with measures of different constructs. As shown in Table 2, in contrast to H16a, the independent self-construal scale correlated more strongly with direct communication style, r = .48, $p < .001$, and coded directness, r = .22, $p < .001$, than with independent TST, r = .06, ns. Similarly, the data were not consistent with H16b. The collective interdependent self-construal scale correlated more strongly with indirect communication style, r = .22, $p < .001$, than with interdependent TST where a negative correlation was observed, r = - .07, ns.

TABLE 5
Correlations and Reliabilities for Revised Scales in Korea

	INDSC	INTSC	RISC	TSTIND	TSTINT	TSTREL	DCS	ICS	SCEN
Independent SC	(.54)								
Interdependent SC	.24**	(.53)							
Relational SC	.07	.38**	(.74)						
TST independent	.12	.01	.03	(.93)					
TST interdependent	-.09	-.02	-.05	-.97**	(.93)				
TST relational	-.08	-.07	.04	-.23**	.01	(.93)			
Direct com. style	.42**	.21**	.09	-.01	.04	.01	(.74)		
Indirect com. style	-.18**	.20**	.13	-.02	.01	-.02	-.43**	(.83)	
Scenario	.21**	-.03	.09	-.02	.01	.09	.07	-.07	(.95)

NOTE: ** $p < .001$

For H16c, scores on the relational interdependent self-construal correlated more strongly with direct communication style, $r = .21$, $p < .001$, than with relational TST scores ($r = .11$, $p < .05$) H16d predicted that direct communication style should correlate more strongly with coded directness ($r = .19$, $p < .001$) than with any type of self-construal. In fact, higher correlations were observed between direct communication style with independent self-construal ($r = .48$, $p < .001$) and relational interdependent self-construal ($r = .21$, $p < .001$). Finally, H16e predicted that indirect communication style would correlate more negatively with coded directness than with any form of self-construal. The results also did not provide support for this claim. Indirect communication style and coded directness had a correlation of $r = -.09$, $p < .01$, while indirect communication style showed a more robust correlation both with independent self construal ($r = -.30$, $p < .001$) and interdependent self-construal ($r = .22$, $p < .001$). These findings showed that self-construal lacked discriminant validity pan-culturally (see Table 2); despite some variation, within-country results also support this conclusion (see Tables 3–5).

Method Effects

Measurement effects were assessed by comparing different construct, different measure correlations with different construct, same measure correlations. The pan-cultural results presented in Table 2 showed that scaled independent self-construal correlated more highly with scaled direct communication style ($r = .48$) than with coded directness ($r = .22$) and scaled

interdependent self-construal correlated more strongly with scaled indirect communication ($r = .22$) than with coded directness ($r = .07$). Similar patterns were observed in the within-country matrices. These findings suggest artifactual methods variance exists for the measures using Likert-type scales. Given that the correlations between open-ended measures all approached zero, no method effects in the coded data are in evidence.

Additional Analyses

The above results were based on revised scales after dropping weak items to obtain unidimensionality. Whereas excluding problematic items should provide a stronger test of validity, critics might argue that dropping items altered the integrity of the original scales, and that the lack of evidence for validity observed above might be a function of discarding items rather than fundamental problems with the scales. In response to this potential criticism, the MMMT analyses were repeated with scales containing all the original items so that it might be ascertained whether or not the dropping of items resulted in substantively different results with regard to convergent, predictive, and discriminant validity. The 16 hypotheses from the MMMT matrix were retested for the scales including all the items. In this re-analysis, the above results were replicated and data were again inconsistent with measurement validity. [14] The self-construal scales and the TST failed the tests of convergent and discriminant validity, and the pattern of correlations was similar to the results with the refined measures. Evidence of method variance was also obtained. Including all the items in each scale therefore did not change the results of this study.

The MMMT analyses were also repeated using a two-factor model for self-construal (i.e., only independent and interdependent self-construal) as originally conceptualized by Markus and Kitayama (1991). The open-ended items on the TST were recoded for independence and interdependence (no relational interdependence) with an intercoder reliability of .90 or higher for the three pairs of coders. Once again, both self-construal scales and the TST failed to show convergent and discriminant validity, and substantial method variance was observed. Correlations for the revised measures of independent self-construal, interdependent self-construal and TST were all near zero.[16] Using the more conventional two-factor model for independent and interdependent self-construal did not improve the results with respect to construct validity. These results suggest that the Singelis self-construal scale, the RISC scale, and the TST are either measuring different constructs, or that the conceptualization of the self-construal construct may need rethinking at the level of theory before valid measurement can be crafted. These possibilities are discussed in the following section.

DISCUSSION

The current research reports a multimethod, multitrait validation study of various measures of self-construal with data collected in the U. S., Japan, and Korea. This research strategy involves crossing constructs with measures to evaluate evidence for convergent, discriminant, and predictive validity. Evidence for convergent validity is established when alternative measures of the same construct are highly and positively correlated and when they function similarly with outside measures. Measures may be said to have disciminant validity to the extent to which they correlate more highly with alternative measures of the same construct than with measures of different constructs. Predictive validity is obtained when measures of different constructs correlate in a manner consistent with extant theory. Finally, when correlations among measures sharing the same method correlate more highly than measures of the same constructs without a common method confound, evidence for method variance is obtained.

The data in the current study are unequivocally inconsistent with the construct validity of the Singelis (1994) self-construal scale, the Cross et al. (2002) RISC scale, and the Kuhn and McPartland (1954) TST. The evidence is strongly inconsistent with convergent validity. The validity coefficients (i.e., same construct, different measure correlations testing convergent validity) ranged from $r = .11$ to $r = -.07$, and all but one were within sampling error of zero. The statistical power for these tests were all greater than .995, so the lack of statistical significance cannot be explained by an inadequate sample size. The upper bound of the 95% confidence interval ($r = .19$) around the single statistically significant validity coefficient ($r = .11 + .08$ for relational self-construal) was far below the minimum ($r = .50$) acceptable value for convergent validity. In both the pan-cultural and the within-country analyses, the validity coefficients were much too small to provide evidence for construct validity. Simply put, in no case did alternative measures of the same construct converge.

Further, in every case, each of the measures of self-construal correlated more highly with a measure of a different construct than with an alternative measure of the same construct. In the most extreme example, scores on the Singelis (1994) independent self-construal scale correlated at $r = .48$ with the direct communication scale but only $r = .05$, with independent self-construal as measured by the TST. These results indicate a complete failure of discriminant validity.

Together, the convergent and discriminant validity results provide strong evidence for the construct invalidity of the various measures of self-construal. Evidence inconsistent with construct validity was observed both in the pan-cultural analyses and the within-country data, for the full

scales and for the revised scales, and for both the two- and three-dimension models of self-construal. Further, nearly identical convergent invalidity results were also reported in Grace and Cramer (2003). Thus, the failure of the self-construal measures to exhibit convergent and discriminant validity is robust.

The failure to demonstrate construct validity with the MMMT matrix suggests that the self-construal and the TST are not measuring the same construct. This might be because one or both of these are invalid measures of self-construal, the different methods assess different and nearly orthogonal aspects of self-construal, or the theory behind independent, interdependent, and relational self-construals is flawed in a way that precludes valid measurement. Which of the nonmutually exclusive explanations accounts for the data is open to debate.

In considering the relative merits of the scales versus the TST, a number considerations are relevant. First, in terms of face validity, the TST seems superior to the scales. The coded self-descriptions appear highly relevant to independent, interdependent, and relational aspects of self-concept while the content of several self-construal scale items seem to reflect alternative constructs (cf. Levine et al., 2003a, 2003b). Second, whereas previous research has found that the TST does not converge with the Singelis (1994) self-construal scales (Grace & Cramer, 2003), some evidence for the construct validity of TST has been obtained in previous multimethod cross-cultural research (e.g., Triandis et al., 1995; Triandis, McCusker, & Hui, 1990). To the current authors' knowledge, the same is not true for the self-construal scales. Third, consistent with Levine et al. (2003a, 2003b), confirmatory factor analyses of the self-construal scales indicated substantial problems with the scales' factor structure. More than half of the Cross et al. (2002) items had to be discarded to obtain fit, and the Singelis (1994) items could not be refined so that the intended two-factor model would provide acceptable fit. These considerations suggest that the scales, in particular, are problematic.

Alternatively, in the current data, the self-construal scales yielded theory-consistent cross-cultural differences to a greater extent than did the TST. Given, however, the results of recent meta-analyses (Levine et al., 2003; Oyserman et al., 2002), these findings are somewhat idiosyncratic. Also, whereas the current data might be interpreted as suggesting that the scales show greater predictive validity compared to the TST, this advantage is less obvious when controlling for common method variance. The scales, however, do show better predictive utility than the TST even when method confounds are removed. Nevertheless, with respect to construct validity, the preponderance of evidence suggests that the scales are probably more suspect than the TST, but validity problems are observable in both types of measures.

It might be argued that, because the various self-construal measures correlated as predicted with measures of direct and indirect communication, the scales should be viewed as having predictive validity. This raises a provocative metatheoretical question concerning if it is reasonable to infer predictive validity from the same findings that provide evidence against discriminant validity. It is our view that from a conceptual or theoretical standpoint (as opposed to a purely applied application) evidence for predictive utility, minus convergent and discriminant validity, is misleading. In the current results, the fact that the self-construal scales correlated more highly with measures of direct communication than with alternative measures of the same types of self-construal precludes the inference that self-construal scales are valid because they correlate with measures of direct communication. This is a case where results that when viewed in isolation are consistent with theoretical predictions, are actually counter to theoretical predictions when viewed in combination with other findings.

These results inform the recent debate over the validity of self-construal scales. Gudykunst and Lee (2003) offered a particularly strong argument in defense of the self-construal scales. Arguing from a nomological network view of construct validity, Gudykunst and Lee (2003) correctly observed that the self-construal scales had correlated in a theory consistent manner with outside measures in over 50 studies. They asserted that this would not be possible if self-construal scales lacked validity. Levine et al. (2003b) responded that the apparent predictive success of self-construal scales might be explained by method variance, spurious effects, and confounded item content. The results from the current study empirically demonstate that it is indeed possible for measures to correlate with certain outside measures as predicted and still be construct invalid. Levine et al., noting that most of the literature used common methods for both the independent and dependent variables (i.e., various scales), speculated that method variance might explain some of the predictive success of the self-construal scales. This finding was supported in the present results. The self-construal scales were found to predict other scales better than coded responses. Levine et al. also speculated that the self-construal scales contained communication items, and thus item content confounds might have contributed to predictive success in previous studies. This, in part, may explain why self-construal scales were more strongly associated with measures of communication style than measures of self-concept.

Correlations can be spurious or reflect confounded measurement; as a result, triangulation with multiple methods and unconfounding constructs and methods with the MMMT approach is necessary for fully testing construct validity. Viewing construct validity solely in terms of Cronbach and Meehl's (1955) nomological network approach is inadequate for the reasons demonstrated here. Nevertheless, the Cronbach and Meehl

approach is more widely known and more widely practiced in many social sciences. More frequent use of the MMMT approach as a supplement to other validation strategies would enhance the quality of measurement in the social sciences.

Whereas the Singelis (1994) scales have been widely used in previous research, the Cross et al. (2000) relational interdependent self-construal scale is relatively new and has been used in few published studies. Contrary to the evidence presented in the original Cross et al. validation study, the current data are largely inconsistent with the validity of the relational self-construal scale. Even though anticipated sex and cultural differences were observed, contrary to Cross et al.'s argument that relational self-construal is more a function of sex differences than cultural differences, cultural differences were more pronounced in the current data.[17] More importantly, several items needed to be deleted to obtain unidimensionality, and the data were inconsistent with convergent and discriminant validity. Bresnahan et al. (2004) also found evidence of problematic items. Together, these data suggest that the Cross et al. scale should not be used to assess relational interdependent self-construal, and that the scale needs refinement.

One potential limitation in the current research design was possible priming by culture artifacts in connection to the use of multiple item measures. Haberstroth, Oyserman, Schwarz, Kuhnen, and Ji (2002) found that participants primed for interdependence (using the pronoun circling task) were more likely to be sensitive to multiple items and consequently avoid redundancy in subsequent answers to the same question compared to participants who were primed for independence. The findings pose an important challenge to the use of multiple items in intercultural research. Because patterns of reliability estimates were not systematically different in different data collection locations, and because Levine et al. (2003a) found that self-construals were impervious to priming inductions, it is unlikely that these artifacts were in operation in the present data. Nevertheless, researchers should be aware of priming artifacts in multiple item measures, especially in cross-cultural research.

CONCLUSION

The current results provide the strongest evidence to date of construct validity problems with self-construal scales. The current study reported a multimethod, multitrait study (Campbell & Fiske, 1959) of three commonly used self-construal measures with data collected in Korea, Japan, and the U.S. The data showed that the Singelis Self-Construal Scale (1994), the Cross, Bacon, and Morris (2000) RISC, and the Kuhn and McPartland (1954) TST lacked convergent and discriminant validity. Substantial method

effects were also observed. These findings held within each of the three countries, in pan-cultural analyses, with the original measures, with weak items deleted, and for both two- and three- factor models of self-construal. Taken in combination with other recent research documenting validity problems (Grace & Cramer, 2003; Levine et al., 2003a, 2003b; Oyserman et al., 2002), the current data suggest that self-construal measurement needs substantial improvement. Perhaps conceptual revision is needed as well.

APPENDIX A

Directions: Read the description given for each of the following situations. Write the exact words that you would say to the offending party in each of these situations.

Situation 1: You have been waiting in line for a long time at an amusement park to get on a ride with your little niece who is whining about the wait. A large party of people shows up and cuts into the line just ahead of you. Would you say something to these people or not? If you would say something, write the exact words that you would say in the box. Otherwise, print the word "Nothing" in the box.

Situation 2: You are in a crowded library just before an important final trying to study. The people sitting next to you are talking loudly. There is no place for you to move and you cannot concentrate. Would you say something to these people or not? If you would say something, write the exact words that you would say in the box. Otherwise, print the word "Nothing" in the box.

Situation 3: You are preparing a group presentation. A student assigned to your group hasn't done anything at all for the project. There is no opportunity for peer evaluation. On the day that your group presentation is scheduled, the student comes to class and sits with your group. Would you say anything to this person or not? If you would say something, write the exact words that you would say in the box. Otherwise, print the word "Nothing" in the box.

Situation 4: You have been driving around for 20 minutes trying to find a parking space. Just then you see someone heading to his/her car. You follow them and then put on the directional signal indicating to others that you are waiting for this space. The car backs out of the space in your direction and just then another vehicle from the opposite direction speeds up and grabs the space you have been waiting for. The person gets out of the car and starts to walk away. Would you say anything to this person or not? If you would say something, write the exact words that you would say in the box. Otherwise, print the word "Nothing" in the box.

APPENDIX B

(Italics indicate items that were retained in the revised scales for the main analysis.)

Directions: Please read over the following definitions.

Direct communication is based on messages that are clearly stated and easily understood by the receiver. While there are many advantages in communicating clearly, sometimes revealing your real opinion can be hurtful to others and create unanticipated relational problems.

Indirect communication is based on messages that are open to interpretation since they only indirectly suggest what is meant. While you may not be understood if you communicate indirectly, using indirect communication can spare the feelings of the other person.

Direct Communication Style

1. I believe that it is important to say exactly what you mean in most situations.

2. I believe that it is generally better to directly say what you mean.

3. The best kind of communication is always clear and precise.

4. Every word used in a message should contribute clearly to what you mean to say.

5. It is unfair to ask a receiver to figure out what a speaker means to say.

6. I believe that direct communication is usually most effective.

7. In most situations, I prefer that others say clearly what they mean.

8. Problems with others should be addressed directly through talk.

9. I believe that messages should be clearly expressed.

10. I believe that you should get to the point in communication.

11. A speaker should communicate clearly.

12. I usually prefer to express my opinions frankly.

13. I prefer clear messages.

Indirect Communication Style

1. In most real-life situations, indirect communication is preferable to direct communication.

2. Indirect communication is generally the more effective form of communication.

3. I often communicate indirectly with other people.

4. It is usually better to hint and let the other person figure out what you mean.

5. The best kind of communication is indirect & open to interpretation.

6. I prefer only to hint at what I need or want.

7. It is generally better to let the other person figure out what you mean.

8. It is generally better to communicate indirectly.

9. I generally like to communicate indirectly with others.

10. Subtle messages are better than those which are very frank.

11. I regularly use ambiguous communication.

12. Sometimes being misunderstood is preferable compared to saying things too directly.

NOTES

1. Items for collective interdependent self-construal emphasize more general group membership (e.g., "I am careful to maintain harmony with my group"), whereas items for relational interdependent self-construal emphasize more committed connections (e.g., " My close relationships are an important reflection of who I am").

2. Oyserman et al. (2001) agreed with Kim and Raja (2003) about the tentative state of self-construal research; unlike Kim and Raja, they cautioned against the uncritical use of self-construal measures.

3. Not all self-construal researchers agree. For example, Kim and Raja (2003) argued against predicting cultural (as defined by nationality) differences in self-construal and against using such hypotheses as evidence for or against validity. Nevertheless, H1–4 reflect expected differences based on the preponderance of the literature.

4. Other scholars (e.g., Kim & Raja, 2003) have claimed that the TST measures a different aspect of self-construal compared to scales. Grace and Cramer (2003) found very little correlation between independent self-construal and TST responses while interdependent self-construal correlated significantly with TST responses. They recommended that future studies use both quantitative and qualitative measures based on a tripartite model of self possibly including relational-interdependent self-construal.

5. Measurement analyses showed that direct and indirect communication items loaded on different dimensions. This necessitated scoring the direct and indirect items separately. Therefore, separate hypotheses for direct and indirect communication styles are advanced so that symmetry exists among hypotheses, method, and results.

6. Sixty people (twenty in each national group) were asked to evaluate the likelihood of occurrence and typicality of each of the four scenarios. The pilot showed that the majority of participants in the U.S. addressed the problem scenarios directly while participants in Korea and Japan were generally more indirect. Even though many people said they would say nothing given the parking scenario, the decision was made to include it in the full study since many participants felt the situation to be realistic and engaging.

7. The Singelis (1994) self-construal scale was selected for two reasons. First, the Singelis scale has published validation studies offering evidence for construct validity (Singelis, 1994). Second, the Singelis scale is the most frequently used self-construal scale in the fields of communication and psychology. Several recent studies have been published using the Singelis scale (e.g., Aaker & Schmidt, 2001; Coon & Kemmelmeier, 2001; Gorski & Young, 2002; Grace & Cramer, 2003; Haberstroh, Oyserman, Schwartz, Kuhnen & Ji, 2002; Hsu, 2002; Norasakkunkit & Kalick, 2002; Rao, Singhal, Ren & Zhang, 2001; Vohs & Heatherton, 2002).

8. Statements coded as personal agency include such descriptions as "I am trusting," "I am happy to be alive," "I am a perfectionist." Group connection is illustrated in statements such as "I am a sorority sister," "I am the oldest child in my family," "I am a Christian." Statements on TST coded as examples of close, committed relationships include "I am in a committed relationship," "I am very close with my family," "I am a loving partner to Matt."

9. Examples of the coding for scenario responses include "The line seems to be getting longer," coded as indirect; "Excuse me, I am trying to study. Could you please be quiet," coded as mitigated direct; and "The end of the line is back there. No cuts!" coded as direct. "Where the hell have you been? We had to do all your work for you and you're getting a big fat zero from us!" was coded as direct but hostile. In the case of a mixed response, the more direct response was coded (e.g., in the case of a message such as "We've been waiting for a long time. Don't be such a pig!" this was coded as hostile direct because of the high face threat attached to such a remark).

10. Confirmatory factor analysis showed that these were two separate unidimensional factors. Direct communication style was unidimensional ($\chi^2 = 10.254$, $df = 9$, $p = .33$) with a standard score coefficient alpha of .82. Indirect communication style was unidimensional ($\chi^2 = 5.190$, $df = 5$, $p = .39$) with a standard score coefficient alpha of .77.

11. There were proportionately more females than males in the sample; therefore, a 2 X 3 (sex by country) ANOVA was conducted. Effects for sex were observed for RISC, scenario response, and TST; however, these effects were trivial (i.e., accounting for less than one percent of the variance).

12. Effect sizes for t-values were obtained by the formula provided by Hunter, Schmidt and Jackson (1982). The observed t-value is divided by the square root of the quantity of t squared plus N – 2. F-values for the omnibus tests are provided in Table 1, and t-values for planned comparisons are available from the first author. Given that (a) sufficient information for subsequent meta-analysis is reported in Table 1, (b) effect sizes are more informative than the values of test statistics, and (c) the current analyses result in a large number of significance tests, estimates of effect size are reported in text and t-values are not reported.

13. In the current study, approximately 15% of all responses in each of the three countries were interdependent. Other cross-cultural studies using TST had similar results. For example, Bond and Cheung (1983) found that less than 20% of responses in Japan, Hong Kong and the U.S. exhibited group reference. Dhawan, Roseman, Naidu, and Rettek (1995), found that 25% of the responses of Indian participants could be classified as group-oriented while 18% of responses for U.S. participants were group-oriented. Rhee, Uleman, Lee, and Roman (1995) found that 21% of Euro-American responses, 23% of Asian-American responses, and 16% of Korean responses were categorized as exhibiting social identity. Watkins et al. (2003) found a similar proportion (25% or less) for small group–allocentric

responses in five countries described as collectivistic (Hong Kong, Nepal, India, Nigeria, and Zimbabwe). The only study showing opposite results was Ma and Shoeneman (1997) who reported that 80% of the responses for Massai and Samburu tribesmen referenced social identity compared to 20% in the U.S.

14. Tables showing correlations for the full scales are available from the first author.

15. Three items in the independent self-construal scale appear to measure communication style rather than self-construal. When these items were removed and the correlation analysis repeated, the correlation with direct communication style decreased from $r = .48$ to $r = .39$.

16. The same finding was true both pan-culturally and within each country. Tables showing correlations for the two-factor data are available from the first author.

17. Females ($M = 3.71$) scored slightly higher than males [$M = 3.56$, $F (1, 560) = 4.39$, $p < .05$, $eta2 < .01$], but the sex effect was smaller than the country effect [$F (2, 560) = 27.46$, $p < .05$, $eta2 = .09$].

REFERENCES

Aaker, J., & Schmidt, B. (2001). Culture-dependent assimilation and differentiation and the self: Preference for consumption symbols in the U.S. and China. *Journal of Cross-Cultural Psychology, 32*, 561–576.

Bond, M. H., & Cheung, T. S. (1983). College students' spontaneous self-concept: The effect of culture among respondents in Hong Kong, Japan, and the United States. *Journal of Cross-Cultural Psychology, 14,* 153–171.

Bresnahan, M. J., Levine, T. R., & Chiu, H. C. (2004). Self-construal as a predictor of communal and exchange orientation in Taiwan and the United States. *Asian Journal of Social Psychology, 7*, 187–203.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait, multi-method matrix. *Psychological Bulletin, 56,* 81–105.

Coon, H. M., & Kemmelmeier, M. (2001). Cultural orientations in the United States: (Re)examining differences among ethnic groups. *Journal of Cross-Cultural Psychology, 32,* 348–364.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Cross, S. E., Bacon, P. L., & Morris, M. L. (2000). The relational-interdependent self-construal and relationships. *Journal of Personality and Social Psychology, 140,* 210–217.

Dhawan, N., Roseman, I. J., Naidu, R. K., & Rettek, S. I. (1995). Self-concepts across two cultures: India and the United States. *Journal of Cross-Cultural Psychology, 26,* 606–621.

Fiske, A. P. (2002). Using individualism and collectivism to compare culture—A critique of the validation and measurement of the constructs: Comment on Oyserman et al. (2002). *Psychological Bulletin, 128,* 78–88.

Gardner, W. L., Gabriel, S., & Lee, A. Y. (1999). "I" value freedom, but "we" value relationships: Self-construal priming mirrors cultural differences in judgment. *Psychological Science, 10,* 321–326.

Gorski, J., & Young, M. A. (2002). Sociotropy/autonomy, self-construals, response style, and gender in adolescents. *Personality and Individual Differences, 32,* 463–478.

Grace, S. L., & Cramer, K. L. (2003). The elusive nature of self-meaurement: The self-construal scale versus the twenty statements test. *The Journal of Social Psychology, 143,* 649–668.

Gudykunst, W. B., & Lee, C. M. (2003). Assessing the validity of self-construal scales: A response to Levine et al. *Human Communication Research, 29,* 253–274.

Gudykunst, W. B., Matsumoto, Y., Ting-Toomey, S., Nishida, T., Kim, K., & Heyman, S. (1996). The influence of cultural individualism–collectivism, self-construals, and individual values on communication styles across cultures. *Human Communication Research, 22,* 510–543.

Haberstroh, S., Oyserman, D., Schwartz, N., Kuhnen, U., & Ji, L. J. (2002). Is the interdependent self more sensitive to question context than the independent self? Self-construal and the observation of conversational norms. *Journal of Experimental Social Psychology, 38*, 323–329.

Hackman, M. Z., Ellis, K., Johnson, C. E., & Staley, C. (1999). Self-construal orientation: Validation of an instrument and a study of the relationship to leadership communication style. *Communication Quarterly, 47*, 183–195.

Hasegawa, T. (1996). *A preliminary study of the use of silence in the United States and Japan*. Paper presented at the Conference on Theory and Research on Communication and Culture, California State University, Fullerton, March, 1996.

Hsu, C. F. (2002). The influence of self-construals, family and teacher communication patterns on communication apprehension among college students in Taiwan. *Communication Reports, 15*, 123–132.

Hunter, J. E., & Gerbing, M. (1982). Unidimensional measurement, second order factor analysis, and causal models. *Research in Organizational Behavior, 4*, 267–320.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Kanagawa, C., Cross, S. E., & Markus, H. R. (2001). "Who am I?" The cultural psychology of conceptual self. *Personality and Social Psychology Bulletin, 27*, 90–103.

Kashima, E. S., & Hardie, E. A. (2000). The development and validation of the relational, individual and collective self aspect (RIC) scale. *Asian Journal of Social Psychology, 3*, 19–48.

Kashima, Y. (2002). Culture and self: A cultural dynamical analysis. In Y. Kashima, M. Foddy, & M. Platow (Eds.), *Self and identity: Personal, social, and symbolic* (pp. 207–226). Mahwah, NJ: Erlbaum.

Kashima, Y., Yamaguchi, S., Kim, U., Choi, S. C., Gelfand, M., & Yuki, M. (1995). Culture, gender, and self: A perspective from individualism-collectivism research *Journal of Personality and Social Psychology, 69*, 925–937.

Kim, M. S. (2002). *Nonwestern perspectives on human communication*. Thousand Oaks, CA: Sage.

Kim, M. S., & Bresnahan, M.I. (1996). Cognitive basis of gender communication: A cross-cultural investigation. *Communication Quarterly, 44*, 53–69.

Kim, M. S., Hunter, J. E., Miyahara, A., Horvath, A., & Bresnahan, M. (1996). Individual vs. culture-level dimensions of individualism and collectivism: Effects on preferred conversational styles. *Communication Monographs, 63*, 29–49.

Kim, M. S., & Raja, N. S. (2003). When validity testing lacks validity. *Human Communication Research, 29,* 275–290.

Kim, M. S., Sharkey, W. F., & Singelis, T. M. (1994). The relationship between individual's self-construals and perceived importance of interactive constraints. *International Journal of Intercultural Relations, 18*, 117–140.

Kim, M. S., Shin, H. C., & Cai, D. (1998). The influence of cultural orientations on the preferred forms of requesting and re-requesting. *Communication Monographs, 65*, 47–66.

Kuhn, M. H., & McPartland, T. S (1954). An empirical investigation of self-attitudes. *American Sociological Review, 19*, 68–76.

Leung, T., & Kim, M. S. (1997). *A revised version of self-construal*. Unpublished manuscript, University of Hawaii.

Levine, T. R., Bresnahan, M. J., Park, H. S., Lapinski, M. K., Wittenbaum, G. M., Shearman, S. M., Lee, S. Y., Chung, D., & Ohashi, R. (2003a). Self-construal scales lack validity. *Human Communication Research, 29*, 210–252.

Levine, T. R., Bresnahan, M. J., Park, H. S., Lapinski, M. K., Lee, T. S., & Lee, D. W. (2003b). The (in)validity of self-construal scales revisited. *Human Communication Research, 29*, 291–308.

Levine, T. R., Lapinski, M. K., Banas, J., Wong, N. C. H., Hu, A. D. S., Endo, K., Baum, K. L., & Anders, L. N. (2002). Self-construal, self and other benefit and the generation of deceptive messages. *Journal of Intercultural Communication Research, 31*, 29–47.

Ma, V., & Schoeneman, T. J. (1997). Individualism versus collectivism: A comparison of Kenyan and American self-concepts. *Basic and Applied Social Psychology, 19*, 261–273.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224–253.

Nelson, G. L., Batal, M. A., & El Bakar, W. (2002). Directness versus indirectness: Egyptian Arabic and U.S. English communication style. *International Journal of Intercultural Relations, 26*, 39–57.

Nomura, N., & Barnlund, D. (1983). Patterns of interpersonal criticism in Japan and the United States. *International Journal of Intercultural Relations, 7*, 1–18.

Norasakkunkit, V., & Kalick, S. M. (2002). Culture, ethnicity, and emotional distress measures: The role of self-construal and self-enhancement. *Journal of Cross-Cultural Psychology, 33*, 56–70.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 1–72.

Park, H. S. (2001). Self-construal as a moderating factor in opinion shifts resulting from exposure to majority opinions. *Communication Reports, 14*, 105–116.

Park, H. S., & Levine, T. R. (1999). The theory of reasoned action and self-construal: Evidence from three cultures. *Communication Monographs, 66*, 199–218.

Rao, N., Singhal, A., Ren, L., & Zhang, J. (2001). Is the Chinese self-construal in transition? *Asian Journal of Communication, 11*, 69–95.

Rhee, E., Uleman, J. S., Lee. H. K., & Roman, R. J. (1995). Spontaneous self-descriptions and ethnic identities in individualistic and collectivistic cultures. *Journal of Personality and Social Psychology, 69*, 142–152.

Sato, T., & Cameron, J. E. (1999). The relationship between collective self-esteem and self-construal in Japan and Canada. *Journal of Social Psychology, 139*, 426–435.

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin, 20,* 580–591.

Sunar, D. (1999). Culture and gender influences on self-concept and the bases of self-esteem: Four Turkish studies. In W. J. Lonner, & D. L. Dinnel (Eds.). *Merging past, present, and future in cross-cultural psychology: Selected papers from the Fourteenth International Congress of the International Association for Cross-Cultural Psychology* (pp. 387–395). Lisse, Netherlands: Swets & Zeitlinger Publishers.

Thompson, C. A., Klopf, D. W., & Ishii, S. (1991). A comparison of social style between Japanese and Americans. *Communication Research Reports, 8*, 165–172.

Triandis, H. C., Chan, D. K. S., Bhawuk, D. P., Iwao, S., & Sinha, J. B. P. (1995). Multimethod probes of allocentrism and idiocentrism. *International Journal of Psychology, 30,* 461–480.

Triandis, H, C, & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology, 74,* 118–128.

Triandis, H. C., Leung, K., Villareal, M. J., & Clack, F. L. (1985). Allocentric versus idiocentric tendencies: Convergent and discriminant validation. *Journal of Research in Personality, 19*, 395–415.

Vohs, K. D., & Heatherton, T. F. (2002). Self-esteem and threat to the self: Implications for self-construals and interpersonal perception. *Journal of Personality and Social Psychology, 81,* 1103–1118.

Watkins, D., Cheng, C., Mpofu, E., Olowu, S., Singh-Senguta, S., & Regmi, M. (2003). Gender differences in self-construal: How generalizable are western findings? *Journal of Social Psychology, 143,* 501–519.

Watkins, D., Yau, J., Dahlin, B., & Wondimu, H. (1997). The Twenty Statements Test: Some measurement issues. *Journal of Cross-Cultural Psychology, 28*, 626–633.