Self-Construal Scales Lack Validity

TIMOTHY R. LEVINE
MARY JIANG BRESNAHAN
HEE SUN PARK
Michigan State University
MARIA KNIGHT LAPINSKI
Western Michigan University
GWEN M. WITTENBAUM
SACHIYO MORINAGA SHEARMAN
SUN YOUNG LEE
DONGHUN CHUNG
Michigan State University
RIE OHASHI
University of the Air

*Self-construal is thought to mediate and explain the effects of culture on a wide variety of outcome variables. A meta-analysis of published cross-cultural self-construal research is reported in this article, and the results across studies suggests that the evidence for the predicted cultural differences is weak, inconsistent, or nonexistent. The results of 3 priming experiments (*N = *121,* N = *99, and* N = *361) suggest that (a) priming does not account for the inconsistent results observed in the meta-analysis, (b) that scores on a self-construal scale appear to be measuring trait-like constructs that are not sensitive to priming, and (c) that measures of self-construals lack convergent validity. The results of several measurement studies (*N = *121, 223, 230, 323, 214, 206, 126, 204, 148, 141, and 150) were inconsistent with the a priori two-factor measurement model in every case. Self-construal scales were found to be radically multidimensional and highly unstable within and across cultures. These results lead us to conclude that catastrophic validity problems exist in research involving the use of self-construal scales in cross-cultural research.*

Timothy R. Levine (Ph.D., Michigan State University, 1992) and *Mary Jiang Bresnahan* (Ph.D., University of Michigan, 1985) are professors in the Department of Communication at Michigan State University, where *Hee Sun Park* (M.A., University of Hawaii, 1998) is an assistant professor. *Maria Knight Lapinski* (Ph.D., Michigan State University, 2000) is an assistant professor in the Department of Communication at Western Michigan University. *Gwen M. Wittenbaum* (Ph.D., Miami University, 1996) is an associate professor at Michigan State University, where *Sachiyo Morinaga Shearman* (M.A., Michigan State University, 2000), *Sun Young Lee* (M.A., Michigan State University, 1999), and *Dong Hun Chung* (M.A., Chung-Ang University, 1999) are doctoral students. *Rie Ohashi* (Ph.D., Michigan State University, 2000) is an associate professor of liberal arts at the University of the Air, Japan. Correspondence should be directed to Tim Levine, Department of Communication, Michigan State University, East Lansing, MI 48824–1212; email: levinet@msu.edu.

Human Communication Research, Vol. 29 No. 2, April 2003 210–252

The idea of self-construals was introduced by Markus and Kitayama in 1991, and since then self-construal has become an increasingly common way of predicting and explaining cultural differences in cognition, emotion, motivation, and communication. Self-construal is thought of as an individual-level cultural orientation and is theorized to mediate and explain the effects of culture on a variety of social behaviors. Several scales exist for assessing self-construal and many authors expound the empirical and theoretical virtues of self-construal. For example, Kim et al. (2000) describe self-construal as "the most useful theoretical framework for integrating findings on cross-cultural research of communication motivation" (p. 263).

Other scholars, however, are skeptical of the validity of self-construal as a cross-cultural explanatory construct. Matsumoto (1999), for example, argues that "examination of the literature that directly tests the assumptions underlying the Markus and Kitayama theory suggests unequivocally that there is little support for them" (p. 306). Similarly, Park and Levine (1999) observe that the results of research examining the link between culture and self-construal appears inconsistent, and they suggest a possible Western bias in the self-construal construct.

Thus, stark disagreement exists over the utility and validity of the self-construal construct. Because a substantial body of research relevant to this controversy already exists, we believe that the conflicting claims of the proponents and critics of self-construal might best be tested with meta-analysis. Meta-analysis allows researchers to look at findings across studies and test for artifactual and random variation in results (Hunter, Schmitt, & Jackson, 1982). The central claims of self-construal research are reviewed in this meta-analysis and four hypotheses central to the validity of self-construal are advanced. Next, the critics' arguments are reviewed and the reasons why self-construal might lack validity are examined. These reasons lead to two additional rival hypotheses. All six hypotheses are then tested against the existing evidence with meta-analysis. Additional priming and measurement studies are reported to explain the results of the meta-analysis. This investigation begins with a brief review of self-construal research.

SELF-CONSTRUAL

Self-construals focus on two aspects of self-concept. Instead of viewing self-concept as a unitary construct, people may be thought of as possessing a number of different selves. These different self-images affect what people "believe about the relationship between the self and others and, especially, the degree to which they see themselves as separate from

others or as connected with others" (Markus & Kitayama, 1991, p. 226). In other words, self-construal is conceptualized as the "constellation of thoughts, feelings, and actions concerning one's relationship to others, and the self as distinct from others" (Singelis, 1994, p. 581). According to Markus and Kitayama (1991), these different images of self influence an individual's cognition, emotion, and motivation.

Markus and Kitayama (1991) propose independent and interdependent construals of the self as an individual-level explanation for culturally-based differences in perception, motivation, and behavior. Self-construal research primarily focuses on how individuals' selves differ across cultures. Socialization differences imposed by different cultures are thought to give rise to different self-concepts, and research on self-construal represents an effort to investigate the relationship between culture and the self. An emphasis is placed on the individual rather than abstract cultural dimensions such as individualism-collectivism. Self-construals are thought to mediate the effects of culture on outcome variables (Gudykunst et al., 1996; Kim et al., 1996). That is, cultural differences result in differences in self-construals, which, in turn, guide motivation, thought, and behavior.

Independent and Interdependent Self-Construals

An independent self-construal is defined as a "bounded, unitary, stable" self that is separate from social context (Singelis, 1994, p. 581). People who emphasize their independent self view themselves as autonomous and invariant across contexts. The independent self-construal is characterized by a view of the self as unique and distinct from others. Individuals are seen as independent people whose behavior stems from internal feelings and thoughts. Expression of one's unique attributes and the pursuit of one's own goals are valued (Markus & Kitayama, 1991; Singelis, 1994).

An interdependent self-construal is defined as a "flexible, variable" self that emphasizes one's connectedness with others (Singelis, 1994, p.581) and "is marked by sensitivity to situations and social contexts" (Kanagawa, Cross, & Markus, 2001, p. 91). The self is defined by one's reference groups (Markus & Kitayama, 1991) and a person's behavior is contingent upon the feelings, thoughts, and behavior of others. Those with an interdependent self-construal are concerned with enacting appropriate behaviors, fostering harmony with others, and with fitting in. Status, roles, relationships, and belongingness are central in the understanding of self. The concept of face is especially important to those with a strong interdependent self-construal (Kim, Sharkey, & Singelis, 1994), and these individuals are more likely to act in accordance with the expectations of others than with their internal wishes or personal attributes (Markus & Kitayama, 1991; Singelis, 1994).

Although these two contrasting views of self are thought to coexist (in varying degrees) within individuals (Gudykunst et al., 1996; Singelis, 1994), the key premise underlying self-construal theory and research is that the relative strength and influence of independent and interdependent self-construals vary systematically among individuals from different cultures. The independent self is argued to be dominant in individuals from the United States and Western Europe, whereas interdependent self-construal is thought to dominate in collectivist non-Western cultures (Gudykunst et al., 1996; Singelis, 1994), especially Japan (Markus & Kitayama, 1991). Thus, self-construals are argued to reflect and explain cross-cultural differences.

A number of self-report scales have been designed to measure self-construals. The three most commonly used scales are those developed by Singelis (1994), Gudykunst et al. (1996), and Leung and Kim (1997). Each of these scales was designed to measure independent and interdependent self-construal as two orthogonal dimensions. These and other self-construal scales have been used to predict a wide variety of outcome variables including: preferred conversational styles (Gudykunst et al., 1996; Kim et al., 1994; Kim et al., 1996; Singelis & Brown, 1995), conflict strategies (Oetzel, 1998a; 1998c; 1999), motivation to comply with others (Park, 2001; Park & Levine, 1999; Park, Levine, & Sharkey, 1998), embarrassment (Sharkey & Singelis, 1995; Singelis, Bond, Sharkey, & Lai, 1999; Singelis & Sharkey, 1995), requesting styles (Kim, Shin, & Cai, 1998), the processing of persuasive messages (Tasaki, Kim, & Miller, 1999), responses to advertising (Wang, Bristol, Mowen, & Chakaborty, 2000), beliefs about learning (Youn, 2000), deception (Lapinski & Levine, 2000; Levine et al., 1999), medical decision making (Kim, Smith, & Yuego, 1999), doctor-patient interaction (Kim et al., 2000), leadership (Hackman, Ellis, Johnson, & Staley, 1999), attribution errors (Krull et al., 1999), perceptions of fairness (Brockner, Chen, Mannix, Leung, & Skarlicki, 2000), anxiety (Kleinknecht, Dinnel, Kleinknecht, Hiruma, & Harada, 1997), self-esteem (Brokner & Chen, 1996; Sato & Cameron, 1999; Singelis et al., 1999; Vohs & Heatherton, 2002), coping (Cross, 1995), self-regulatory processes (Lee, Aaker, & Gardner, 2000), and the use of self- and other-promoting statements (Ellis & Wittenbaum, 2000). An examination of the research cited above suggests that self-construals have become an increasingly popular way of studying cultural differences.

Hypotheses Central to the Validity of Self-Construals

As noted previously, self-construals are conceptualized as individual-level culture, and it is argued that they mediate and explain cross-cultural differences. Consequently, the validity of self-construal theory and research rests, at minimum, on the existence of systematic cultural differ-

ences in the relative strengths of the two types of self-construals. That is, for self-construals to mediate and explain cultural differences, self-construals must reflect the intended cultural differences. More precisely, the theoretical and construct validity of self-construals require (a) that the independent self-construal is dominant in individuals from Western cultures and (b) that the interdependent self-construal is more pronounced in non-Western cultures, especially Asian cultures. These predicted cultural differences lead to four hypotheses central to the validity and utility of self-construals. These four hypotheses are as follows:

H1: Individuals from Western cultures will score higher on measures of independent self-construal than will individuals from Asian cultures.
H2: Individuals from Asian cultures will score higher on measures of interdependent self-construal than will individuals from Western cultures.
H3: Individuals from Western cultures will score more higher on measures of independent self-construal than on measures of interdependent self-construal.
H4: Individuals from Asian cultures will score higher on measures of interdependent self-construal than on measures of independent self-construal.

Criticisms of Self-Construals

The sheer quantity of data involving self-construals has increased dramatically over a relatively short period of time, and many published research articles claim to document the validity or utility of self-reported measures of self-construals (e.g., Gudykunst et al., 1996; Hackman et al., 1999; Kim et al., 2000; Singelis, 1994; Singelis et al., 1999). Even when the data might be interpreted as inconsistent with the validity of self-construals, authors seem reluctant to criticize the construct (e.g., see Gardner, Gabriel, & Lee, 1999; Gudykunst et al., 1996; Hackman et al., 1999; Kanagawa, Cross, & Markus, 2001; Kim et al., 1996; Oyserman, Coon, & Kemmelmeier, 2002; Sato & Cameron, 1999). In perhaps the most dramatic example, even though Gudykunst et al. found that the "means [on self-construals] suggest that the samples do not reflect the general cultural tendencies usually associated with the four cultures," they concluded that:

> the present data support the hypotheses that independent self-construals and individualistic values mediate the influence of cultural I-C [individualism-collectivism] on LC [low context] communication, and that interdependent self-construals and collectivistic values mediate the influence of cultural I-C on HC [high context] communication. The results further suggest that self-construals and values are better predictors of and account for more variance in LC and HC communication styles than does cultural I-C. (1996, p. 530)

In short, when their self-reported measures of self-construals failed

to show expected cultural differences, Gudykunst et al. concluded that self-reports tell us more about cultural differences than objective national identity.

Self-construals, however, have not been completely immune to criticism. Matsumoto's (1999) recent review of the literature led him to conclude that the existing data are not consistent with the validity of self-construals. Matsumoto argued (a) that Markus and Kitayama's (1991) conclusions rest on the assumption that there are systematic cultural differences in self-construals and (b) that the data Markus and Kitayama drew upon in their seminal article failed to test this assumption. Matsumoto further argued, citing Gudykunst et al. (1996) and Kim et al. (1996) among others, that subsequent research testing cultural differences in self-construals does not support Markus and Kitayama's claims. Matsumoto concluded that "the evidence available to date severely challenges the validity of their theoretical framework for explaining observed national differences in psychological phenomena" (p. 301). Park and Levine (1999) independently drew similar conclusions.

Rival Hypotheses and the Reasons Why Self-Construals Might Fail

A careful examination of the literature reveals at least four reasons why we might expect the data to be inconsistent with the four hypotheses central to the validity of self-construals. These reasons include (a) sensitivity to situational priming; (b) the existence of a Western, independence bias; (c) faulty scale construction and validation; and (d) an overly simplistic conceptualization of self-construal. Because the meta-analysis can only address the first two of these reasons, these two are the initial focus here.

Gardner et al. (1999) attempted to experimentally manipulate self-construals, independent of culture, with priming methods. Participants were primed by reading independent or interdependent stories or by doing a word search for pronouns ("I" and "me" or "we" and "us"). Gardner et al. found that the balance between individualist and collectivist values was shifted by the priming task. Further, the priming manipulations were influential enough to shift general cultural tendencies. Participants in the U.S. who were primed for interdependence scored higher on collectivist values than individualist values. The reverse was true for participants in Hong Kong. These results and others (e.g., Brewer & Gardner, 1996; Kuhen & Hannover, 2000; Trafimow, Silverman, Fan, & Law, 1997; Trafimow, Triandis, & Goto, 1991; Ybarra & Trafimow, 1998) suggest that self-construals are highly sensitive to situational priming.

Although Gardner et al. (1999) did not directly measure self-construals with a self-construal scale, other studies have shown that priming affects responses to open-ended measures of self-construal (e.g., Brewer &

Gardner, 1996; Trafimow et al., 1991; Trafimow et al., 1997). Thus, it is possible (if not likely) that priming could affect scores on commonly used self-construal scales. Because the interdependent self-construal is conceptually defined as flexible and variable, it should be especially sensitive to priming. If this is the case, it might be possible for self-construal researchers to inadvertently prime research participants in some part of the data collection procedures. Or, even if the researcher managed complete control over the research procedures so that no priming occurred, some external event might prime the participants. In short, it is plausible that scores on self-construal scales might often reflect situational priming rather than (or in addition to) stable individual-level cultural tendencies.

If situational priming has contaminated previous self-construal research, we believe that this would be reflected in the extant data. Specifically, it is reasonable to expect that such priming, if it occurred, would vary widely from study to study. This would instill substantial variability in results from study to study that is not explainable in terms of sampling error or identifiable moderators. In short, to the extent that situational priming affects self-construal findings, one might expect radical, and apparently inexplicable, inconsistencies in results from study to study. This line of reasoning allows us to posit our first rival hypothesis which we call the *situational priming hypothesis*.

> H5: There will be significant and substantial variability in the effect sizes relevant to each of the four self-construal hypotheses.

A second possible problem with self-construals concern allegations of a Western bias in the scales. Park and Levine (1999) speculated that

> it is also possible that the theoretical separation of independent self-construal and interdependent self-construal concept at the individual level is a Western concept. The members of Eastern, or collectivistic, cultures may construe their self-images as independent and/or interdependent with others depending on situation or task types (p. 215).

Other researchers have voiced concern over a Western bias in the specific measurement approach used in self-construal scales (Fiske, 2002; Kanagawa et al., 2001; Kitayama, 2002; Markus & Kitayama, 1998). Kanagawa et al. argued that widely used structured questionnaires may be culturally biased and ill suited to studying the dynamic nature of the interdependent self-construal. Similarly, Markus and Kitayama suggest that self-report scales "are most appropriate for those with a Euro-American personality" (p. 75) and may be ill suited for use in Asian cultures. To the extent that this speculation has merit, we would expect self-construal

scales to work better in Western cultures than in Asian cultures. This expectation leads to our second rival hypothesis, the *Western bias hypothesis*. This hypothesis has two parts:

H6a: The separation between the two types of self-construals will be evident in data from Western cultures (in accordance with H3) but not in data from Asia (contrary to H4).

H6b: More generally, the effects for independent self-construal will be larger than the effects for interdependent self-construal, and the effects from data collected in Western cultures will be larger than results from Asian cultures.

META-ANALYSIS

Method

Study selection. This meta-analysis tests the four hypotheses central to self-construals as well as the two rival hypotheses. In selecting studies for the meta-analysis, five inclusion criteria were specified a priori. First, only studies reporting results in necessary detail were included, meaning they reported either (a) self-construal means and standard deviations or (b) significance tests for the relevant comparisons. Studies by Redford (1998) and Youn (2000) were excluded because the results were not reported in sufficient detail. Second, only published studies were analyzed. Third, only studies using one of the three primary self-construal scales (i.e., Gudykunst, 1996; Leung & Kim, 1997; Singelis, 1994) or their close variants were considered. Fourth, the meta-analysis was limited to studies that reported actual cross-cultural data from individuals in different national cultures. This excluded studies comparing the self-construals of participants with different ethnicity in which the data was collected in a single location or country (e.g., Lapinski & Levine, 2000; Oetzel, 1998b; Singelis, 1994; Singelis & Brown, 1995; Singelis & Sharkey, 1995). Finally, data from Hawaii were also excluded.

The first inclusion criteria is obviously essential because there must be some basis for calculating effect sizes in order to cumulate the results. The other four inclusion criteria, however, involved more difficult decisions and may be more controversial. As a rule, we selected the criteria in order to give self-construals the fairest test possible. For example, the quality of the original research is an issue in meta-analysis. We only included published studies under the assumption that research published in refereed journals should be of an overall higher quality than unpublished work. Further, to the extent that differences exist between the results of published and unpublished work, it is more likely that self-construals would receive support in published work because editorial practices may have a confirmation bias (Meehl, 1986). Studies (e.g., Brockner & Chen, 1996) using self-construal scales that treated interde-

pendence and independence as opposite ends of the same dimension were excluded because they are clearly measuring the construct(s) differently than the more commonly accepted scales. Cross's (1995) data were excluded because she used ego-task and collectivism scales to measure self-construals. Research testing for differences in self-construals among people of different ethnicity but residing in the same location (e.g., Okazaki, 2000; Oetzel, 1998b) were excluded because, if self-construals are valid, stronger effects should be observed in truly cross-cultural data. Finally, data from Hawaii is excluded because Hawaii is usually predicted to fall between the West and Asia, which should dilute effects. In short, the criteria were chosen in such a way as to maximize the chances that the self-construal hypotheses would receive support if they are, in fact, true. Similarly, the criteria employed should provide a more rigorous test of the rival hypotheses.

Procedures and analyses. The initial step was to obtain all available findings that met the five inclusion criteria. Relevant studies were obtained from the authors' files and through computer searches using *ProQuest* and *PsycINFO*. The reference sections of all obtained studies were examined to find citations to other relevant studies. These procedures led us to identify eight studies which met the criteria (see Table 1).

After the studies were obtained, usable and relevant effect sizes were calculated. In most cases, *t* values were calculated based on means, standard deviations, and sample sizes. The *t* values were then converted to *r*. Krull et al. (1999) did not provide standard deviations so *r* was calculated from *F*.

The unit of analysis was the effect under scrutiny rather than the study (cf. Oyserman et al., 2002). So, for example, Gudykunst et al. (1996) yielded 4 effects relevant to H1. In all, 13 effects (N = 4,527) relevant to Hypothesis 1, 13 effects (N = 4,669) for H2, 8 effects (N = 3,144) for H3, and 9 effects (N = 3,684) for H4 were analyzed. The Hunter approach to meta-analysis was used (Hunter, Schmidt, & Jackson, 1982). Weighted (by sample size) mean effect sizes (r) were calculated for the first four hypotheses, along with chi-square tests for homogeneity of effects.

Results

Hypotheses 1 though 4. Hypothesis 1 predicted that individuals from Western cultures would score higher on measures of independent construal than individuals from Asian cultures. Thirteen usable tests of this hypothesis were found in the literature, with a total N = 4,527. Seven of those 13 effects (54%) were statistically significant in the predicted direction, and there were no significant findings in the wrong direction. The observed effect sizes (in r) ranged from -.075 to +.528. Consistent with H1, the mean effect was statistically significant in the predicted direction,

TABLE 1
Meta-Analysis of Cultural Differences in Self-Construal

Study	*Scale*	*Location(s)*	*Effect* (r)	N
H1: Western > Asian on independent self-construal[a]				
Gudykunst et al. (1996)	Gudykunst	U.S.-Japan	+.034	475
	Gudykunst	U.S.-Korea	+.145*	451
	Gudykunst	Australia-Japan	-.075	302
	Gudykunst	Australia-Korea	+.036	278
Kim et al. (2000)	Leung & Kim	U.S.-Hong Kong	+.528*	506
Kim et al. (1996)	Kim	U.S.-Japan	+.210*	474
	Kim	U.S.-Korea	+.430*	438
Kleiknecht et al. (1997)	Singelis	U.S.-Japan	.000	342
Krull et al. (1999)	Singelis	U.S.-Taiwan	+.119	96
	Gudykunst	U.S.-PRC	+.431*	76
Park & Levine (1999)	Leung & Kim	U.S.-Korea	+.202*	294
Sato & Cameron (1999)	Singelis	Canada-Japan	+.020	292
Singelis et al. (1999)	Singelis	U.S.-Hong Kong	+.400*	503
H2: Asian > Western on interdependent self-contrual[b]				
Gudykunst et al. (1996)	Gudykunst	Japan-U.S.	+.019	475
	Gudykunst	Japan-Australia	-.053	302
	Gudykunst	Korea-U.S.	+.206*	551
	Gudykunst	Korea-Australia	+.126	278
Kim et al. (2000)	Leung & Kim	Hong Kong-U.S.	+.329*	505
Kim et al. (1996)	Kim	Japan-U.S.	-.088	524
	Kim	Korea-U.S.	+.006	433
Kleiknecht et al. (1997)	Singelis	Japan-U.S.	-.142*	342
Krull et al. (1999)	Singelis	Taiwan-U.S.	+.322*	96
	Gudykunst	PRC-U.S.	+.288*	76
Park & Levine (1999)	Leung & Kim	Korea-U.S.	+.603*	292
Sato & Cameron (1999)	Singelis	Japan-Canada	-.334*	292
Singelis et al. (1999)	Singelis	Hong Kong-U.S.	+.299*	503
H3: Independent > interdependent self-construal in Western cultures[c]				
Gudykunst et al. (1996)	Gudykunst	U.S.	+.618*	566
	Gudykunst	Australia	+.502*	220
Kim et al. (2000)	Kim	U.S.	+.767*	419
Kim et al. (1996)	Leung & Kim	U.S.	+.177*	473
Kleiknecht et al. (1997)	Singelis	U.S.	+.103*	362
Park & Levine (1999)	Leung & Kim	U.S.	+.787*	300
Sato & Cameron (1999)	Singelis	Canada	+.227*	344
Singelis et al. (1999)	Singelis	U.S.	+.239*	464

TABLE 1 Continued
Meta-Analysis of Cultural Differences in Self-Construal

Study	*Scale*	*Location(s)*	*Effect (r)*	N
H4: Interdependent > Independent Self-Construal in Asia[d]				
Gudykunst et al. (1996)	Gudykunst	Japan	-.544*	384
	Gudykunst	Korea	-.390*	336
Kim et al. (2000)	Yeung & Kim	Hong Kong	-.241*	592
Kim et al. (1996)	Kim	Japan	-.055	580
	Kim	Korea	+.301*	398
Kleiknecht et al. (1997)	Singelis	Japan	-.241*	322
Park et al. (1999)	Yeung & Kim	Korea	-.131*	290
Sato & Cameron (1999)	Singelis	Japan	-.485*	240
Singelis et al. (1999)	Singelis	Hong Kong	+.451*	542

NOTE: Positive effects are in the predicted direction. Negative effects indicate effects in the wrong direction. [a] Weighted mean effect $r = .2064 \pm .106$, $p < .05$, $N = 4{,}527$, $K = 13$; homogeneity of effects, $\chi^2(12) = 185.14$, $p < .001$; percent of variance in effects attributable to sampling error = 7.02%; residual variance in effects not attributable to sampling error = 92.98%. [b] Weighted mean effect $r = .1065 \pm .1235$, $p = ns$, $N = 4{,}669$, $K = 13$; homogeneity of effects, $\chi^2(12) = 246.35$, $p < .001$; percent of variance in effects attributable to sampling error = 5.28%; residual variance in effects not attributable to sampling error = 94.72%. [c] Weighted mean effect $r = .4268 \pm .174$, $p < .05$, $N = 3{,}144$, $K = 8$; homogeneity of effects, $\chi^2(7) = 298.81$, $p < .001$; percent of variance in effects attributable to sampling error = 2.68%; residual variance in effects not attributable to sampling error = 97.32%. [d] Weighted mean effect $r = -.1038 \pm .212$, $p = ns$, $N = 3{,}684$, $K = 9$; homogeneity of effects, $\chi^2(8) = 395.17$, $p < .001$; percent of variance in effects attributable to sampling error = 2.28%; residual variance in effects not attributable to sampling error = 97.72%.
*$p < .05$.

$r = .206$, with a confidence interval of $\pm .106$. However, the mean effect was rather small, and the effects were significantly and substantially heterogeneous, $\chi^2(12) = 185.14$, $p < .001$. Only 7.02% of the observed cross-study variance in effects sizes was attributable to sampling error, meaning that 92.98% of the variance was caused by unknown factors. This variability in effects can be clearly seen in Table 1 where more than half of the observed effects (54%) fell outside the 95% confidence intervals around the mean correlation. This suggests the presence of one or more powerful moderators and that the cross-study mean correlation should not be considered an estimate of a population value (cf. Oyserman et al., 2002).

The second hypothesis predicted that individuals from Asian cultures would score higher on measures of interdependent construal than individuals from Western cultures. Again, 13 usable tests of this hypothesis

were analyzed (total N = 4,669). Six of those 13 effects (46%) were statistically significant in the predicted direction, but two were significant in the wrong direction, with the observed effect sizes ranging from -.334 to +.603. Inconsistent with H2, the mean effect was not statistically significant, but it was in the predicted direction (r = .107). The 95% confidence interval was ±.1235. Again, the observed effects were significantly and substantially heterogeneous, $\chi^2(12) = 246.35$, $p < .001$, with 94.72% of the variance in effects not due to sampling error. For H2, a clear majority of the findings from individual studies (70%) fell outside the 95% confidence intervals around the mean correlation. Thus, the data clearly do not support H2.

Individuals from Western cultures were predicted to score higher on measures of independent construal than on measures of interdependent self-construal according to H3. Eight effects (N = 3,144) relevant to H3 were assessed. Here, each individual effect was statistically significant in the predicted direction, with the observed effect sizes ranging from +.103 to +.787. Consistent with H3, the mean effect was statistically significant, substantial, and in the predicted direction (r = .427) with a relatively wide confidence interval of ±.174. However, once again, significant and substantial heterogeneity was evident, $\chi^2(7) = 298.81$, $p < .001$. Only 2.68% of the observed cross-study variance in effects sizes was attributable to sampling error.

Finally, H4 specified that individuals from Asian cultures will score higher on measures of interdependent construal than on measures of independent self-construal. Nine usable effects testing the hypothesis were examined, with a total N = 3,684. Six of those 9 effects (66%) were statistically significant in the wrong direction, whereas only two (22%) were significant in the predicted direction. The observed effect sizes varied dramatically from -.544 to +.451. Inconsistent with H4, the mean effect was neither statistically significant nor in the predicted direction (r = -.104) with a confidence interval of ±.212. The effects were massively heterogeneous, $\chi^2(8) = 395.17$, $p < .001$, with a mere 2.28% of the observed cross-study variance in effects sizes attributable to sampling error. Thus, the extant data are clearly inconsistent with H4.

Rival hypotheses 5 and 6. The situational priming hypothesis (H5) predicted that there would be significant and substantial variability in the effect sizes relevant to Hypotheses 1 through 4. The data were unequivocally consistent with this prediction. Statistically significant and substantial variability in effects were reported for each of the first four hypotheses (see the chi-square tests above). Residual variances in individual findings not attributable to sampling error ranged from 92.98% for H1 to 97.72% for H4. Thus, consistent with the situational priming hypothesis, the literature is characterized by radically inconsistent findings. Obviously, one or more powerful moderators of artifacts are in operation.

The first Western bias hypothesis (H6a) proposed that the separation between the two types of self-construals would be evident in data from Western cultures (as predicted by H3) but not in data from Asia (inconsistent H4). The data were consistent with this hypothesis as well. Every study in this investigation found significant differences between the two self-construals in Western cultures with scores on independence being greater than scores on interdependence (mean effect +.43). The mean effect for H4, specifying that Asians would score higher on interdependent, rather than independent self-construals, was small and inconsistent (-.10).

Because the data were more consistent with H1 than H2 and with H3 than H4, the data were also more consistent with H6b. Measures of independent self-construal seem to work better than interdependent self-constual and the measures of self-construals seem to work better in the West than in Asia. Thus, a Western bias remains plausible although support for H6a and H6b is qualified by support for H5. The results of the meta-analysis are summarized in Table 1.

Discussion

Four hypotheses central to the validity of self-construals and two rival hypotheses were tested with meta-analysis. The one inescapable conclusion drawn from the existing data is that self-construal results are radically inconsistent across studies. Simply put, the findings are all over the place.

Little support was found for the predicted effects of culture on self-construals. The data were clearly inconsistent with two of the self-construal hypotheses. Across studies and more than 4,000 subjects, the evidence does not support the claim that those in Japan, Korea, China, or Taiwan are more interdependent than those living in the U.S., Canada, or Australia. Similarly, the data clearly contradict the claim that those in Asian countries are more interdependent than independent.

The results, however, provide superficial support for the other two self-construal hypotheses in so far as the mean correlations are significantly different from zero and in the predicted direction. In one case, the mean effect was small and inconsistent, whereas in the other, the mean effect was moderate to large and inconsistent. Because the findings were so heterogeneous, however, the individual effects are not additive and the across-study mean effect has little, if any, substantive meaning. Specifically, if the findings from individual studies provided valid tests of the same hypothesis, then the mean correlation would be an estimate of the population correlation (ρ) and the sample correlations would be (approximately) normally distributed around the mean correlation with the vast majority of the findings from individual studies falling within the

95% confidence intervals. The extreme variability in the observed individual effects leads us to reject the idea (with a high degree of confidence, $p < .001$) that the sample correlations come from the same population. Thus, the effects are not additive and the mean correlations cannot be meaningfully interpreted as supporting the hypotheses. Instead, the data are more accurately interpreted as indicating the presence of one or more powerful moderators, artifacts, or confounds.

Two identifiable potential moderators are the different self-construal scales and the particular countries where the data was collected. It is possible that one scale might be superior to another or that predicted differences might be more evident in one country than another. Unfortunately, the number of effects for each hypothesis is too small to formally include these variables in the meta-analysis. However, scale and country are listed in Table 1, and some feel for the plausibility of these factors as moderators can be gained through visual inspection. Careful examination of the results suggests that the inconsistencies are not solely attributable to scale or country. Clearly, researchers using the same scales find inconsistent results. Similarly, inconsistent results are readily observed between studies making the same country comparisons. Thus, we do not believe that differences in the measures used or the countries tested can adequately account for the heterogeneity of effects.

Considering the evidence for the four self-construal hypotheses together, the meta-analysis reported here provides statistical evidence for Matsumoto's (1999) claim that the existing data "severely challenges the validity" of self-construals as individual-level culture orientation and as a mediator of cultural differences (p. 289). The cumulative results across existing studies suggest that the evidence for predicted cultural differences in self-construal is, depending on the particular hypothesis, weak, inconsistent, or nonexistent. For example, of the eight studies analyzed here, only one (Singelis et al., 1999) provides statistically significant support for all four hypotheses. But, the Singelis et al. (1999) results fall outside the 95% confidence intervals on all four hypotheses making those supportive results statistical outliers. In short, the data are much more consistent with the observations of those critical of self-construals than with the advocates of self-construals.

The data are consistent with Park and Levine's (1999), Kanagawa et al.'s (2001), and Markus & Kitayama's (1998) speculation concerning a Western bias. The results indicate that independent self-construal seems to work better than interdependent self-construal and that the two types of self-construals seem to be more differentiated in participants from Western cultures than in participants from Asian cultures.

One might take exception with one or more of our inclusion criteria and argue that had we included additional studies, then the data would

surely have been more consistent with the validity of self-construals. We doubt that such arguments would be convincing for two reasons. First, although we did use rather limiting inclusion criteria that excluded a potentially large number of studies, the inclusion criteria should (theoretically, at least) increase rather than decrease the likelihood that the meta-analysis would be consistent with the validity of self-construals. For example, if self-construals are valid, we would be more likely to see support in published rather than unpublished studies and in studies comparing the West to Asia rather than Hawaii to the mainland or Hawaii to Asia. Stated differently, had we thrown a wider net and included more studies, we would expect less rather than more support for the validity of self-construals and more rather than fewer inconsistencies in the results of individual studies. Arguments to the contrary are likely to run into logical inconsistencies.

Second, and more importantly, given the nature of our findings, it is difficult realistically to imagine how the inclusion of additional studies could change the findings much or make them much more supportive of the validity of self-construals. For the sake of argument, let's suppose that we overlooked some studies in which the data were perfectly consistent with all four self-construal hypotheses. Although the inclusion of such studies would move the mean effects in the direction of the self-construal predictions, it would also increase the variance in observed effects making the findings even more inconsistent than they already are. Because the primary conclusion reached from the current data is that the results are radically inconsistent, the inclusion of additional studies is unlikely to change this.

There is one class of excluded studies, however, that deserves special attention. Six studies were excluded because the data were collected in a single location and hence these were not deemed true cross-cultural studies. Lapinski and Levine (2000), Singelis (1994), Singelis and Brown (1995), and Singelis and Sharkey (1995) compared participants of Asian descent to those of Caucasian ancestry in Hawaii. Oetzel (1998b) and Okazaki (2000) compared American and International students in the U.S. Theoretically, we would expect such studies to provide a weaker test of the self-construal hypotheses. However, an informal examination of the results from these studies show that five out of the six, including all four studies done in Hawaii, yield data consistent with the self-construal hypotheses. Thus, had these studies been included in the meta-analysis, it is likely that the mean effects would move slightly in a direction more consistent with Hypotheses 1 though 4. This, however, also raises the question of why self-construals appear to work better in Hawaii than in true cross-cultural comparisons. Theoretically, if self-construals are valid, the results should be the other way around. These findings be-

come understandable if the argument for self-report scales being more valid in data from the United States than elsewhere has merit (e.g., Fiske, 2002; Kanagawa et al., 2001; Kitayama, 2002; Markus & Kitayama, 1998).

PRIMING STUDIES

Given the failure of self-construals in previous research, one might question why self-construals fail and why the data appear so dramatically inconsistent from study to study. The meta-analysis undoubtedly demonstrates the existence of one or more strong and unidentifiable moderators, confounds, or artifacts. That is, something is causing self-construal results to vary radically from study to study. Perhaps the most plausible account for the radical inconsistency in effects is situational priming. If self-report measures of self-construals are highly sensitive to priming and if priming is assumed to vary from study to study, then we would predict exactly the sort of variability in effects indicated by the meta-analysis.

Theoretically, interdependent self-construal should be highly sensitive to situational priming. Interdependent self-construal is defined conceptually as flexible and variable (Singelis, 1994) and is "marked by sensitivity to situations and social contexts" (Kanagawa et al., 2001, p. 91). Thus, the malleable nature of interdependent self-construal should make it highly responsive to situational priming. Alternatively, independent self-construal is conceptualized as a stable self that is separate from, and relatively invariant across, social contexts (Kanagawa et al., 2001; Markus & Kitayama, 1991; Singelis, 1994). The stable nature of independent self-construal should make it relatively more resistant to priming effects. Consistent with this reasoning, Kanagawa et al. found that self-descriptions of Japanese students varied more across situations than did American student's self-descriptions. Thus, to the extent that self-report, self-construal scales reflect the constructs they are intended to measure, we advance the first of three rival priming hypotheses:

H7a: Situational priming will affect interdependent but not independent self-construal.

The results of previous priming research and our meta-analysis, however, suggest a different prediction. Priming studies have found that priming effected both types of self-construals and that priming effects are evident for both participants in the U.S. and participants in non-Western countries (Brewer & Gardner, 1996; Gardner et al., 1999; Kuhen & Hannover, 2000; Trafimow et al., 1991; Trafimow et al., 1997; Ybarra & Trafimow, 1998). The meta-analysis found massive variability in the re-

sults of both types of self-construals. These data lead to the following alternative prediction:

H7b: Situational priming will affect both interdependent and independent self-construals.

A third possibility is suggested by Kanagawa et al. (2001). They imply that structured scales, like each self-construal scale, may not be well-suited for studying the dynamic aspects of self-concept. To the extent that self-construal scales tap stable, trait-like aspects of self-concept and fail to capture the dynamic aspects of the self, we advance a third possibility:

H7c: Situational priming will have no effect on interdependent and independent self-construals.[1]

Although the results of the meta-analysis were most consistent with H7b, priming was not manipulated in any study under analysis. Therefore, there is no direct evidence from the current results to show that priming caused the between-study variance in effects. To directly test the three rival priming hypotheses and in an attempt to explain the results of the meta-analysis, we conducted three priming experiments.

Priming Study 1

Method

The participants were 121 (38 male and 83 female) undergraduate students at a large Midwestern U.S. university (mean age = 21.46, *SD* = 1.23). Most participants were Caucasian (84.3%) and all participants received extra course credit in exchange for their participation.

The study used a posttest only experimental design with two independent groups and random assignment. The independent variable was prime type ("I" or "we") and dependent measures were average scores on independent and interdependent self-construal scales.

The priming induction and instructions were identical to the pronoun search task used by Gardner et al. (2000).[2] Participants were asked to read a paragraph and circle all the pronouns found within. The pronouns were either first-person singular ("I," "my," "me") or first-person plural ("we," "our," "us"). Twenty pronouns were embedded within the otherwise identical paragraphs. The number of pronouns identified was checked to ensure that respondents completed the priming task.

Immediately following the pronoun search, the participants completed Leung and Kim's (1997) self-construal scale. This scale included 29 (14

interdependent and 15 independent) Likert-type items with 5-point response formats. Scale reliabilities were acceptable (interdependence α = .70 and independence α = .86).

Results and Discussion

On independent self-construal, the mean item score for those in the I-prime condition was 4.29 and the mean in the we-prime condition was 4.19. The means, although in the direction predicted by H7b, were not significantly different, $t(119) = 1.43$, $p = .156$, $\eta^2 = .016$. Contrary to both H7a and H7b, little difference between priming conditions was observed on mean interdependence scores; I-prime $M = 3.09$, we-prime $M = 3.12$, $t(119) = 0.41$, $p = .682$, $\eta^2 = .00$. The statistical power was .29 for small effects ($r = .10$) and .86 for medium effects ($r = .24$). The data were reanalyzed, splitting the data by self-identified ethnicity, but all results remained statistically nonsignificant. Thus, the data were inconsistent with both H7a and H7b.

Although the data were consistent with H7c, which predicted no differences, a lack of statistical power and methodological limitations might explain the results. There was a nonsignificant trend for independent self-construal. One might speculate on whether or not that trend would have been statistically significant if the sample size had been larger. Further, only one type of priming manipulation was used. Therefore, additional data is needed before H7c can be reasonably accepted. In Priming Study 2, an additional type of prime was used.

Priming Study 2

Method

The participants were 99 (34 male and 65 female) undergraduate students at the same large Midwestern U.S. university. All participants were Caucasian. Participants received course credit in exchange for their participation.

The design was a 2 (story, pronoun search) by 2 (independent prime, interdependent prime) independent groups design with mean scores on the independent and interdependent self-construal scales as the dependent measures. The pronoun priming induction and instructions were identical to the priming task used in Gardner et al. (2000) and Priming Study 1. The story prime has been successfully used in several previous priming studies (e.g., Gardner et al., 1999; Trafimow et al., 1991; Trafimow et al., 1997; Ybarra & Trafimow, 1998) and involves the story of a warrior-king who must decide on a commander of a detachment of troops. His

selection is either based on the potential for personal gain (independent prime) or family loyalty (interdependent prime). Immediately following the pronoun search, the participants completed Leung and Kim's (1997) self-construal scale. Scale reliabilities were consistent with Study 1 (interdependence $\alpha = .75$ and independence $\alpha = .91$).

Results and Discussion

The type of priming induction (story or pronoun search) had negligible effects (Fs < 1.00) on the self-construal scales and did not interact with the type of self-construal primed ($F \leq 1.50$). Therefore, hypotheses 7a, 7b, and 7c were tested by averaging across priming type.

On independent self-construal, the mean item score for those in the independent prime condition was 4.17 and the mean in the interdependent-prime condition was 4.13. These means were not significantly different, $F(1, 89) = 0.09$, $p = ns$, $\eta^2 = .00$. No difference between priming conditions was observed on interdependence scores either; independent prime $M = 3.18$, interdependent prime $M = 3.05$, $F(1, 90) = 1.36$, $p = ns$, $\eta^2 = .01$. The statistical power was .23 for small effects ($r = .10$) and .75 for medium effects ($r = .24$). Thus, the data were again inconsistent with both H7a and H7b.

It is possible that the lack of priming effects for both types of self-construals might be attributable to the ordering of the self-construal items. Because respondents in both Studies 1 and 2 completed both types of self-construal items, answering previous items could prime that particular self-construal, diluting the priming induction. Also, statistical power in Study 2 was again less than ideal. For these reasons, a third replication was conducted.

Priming Study 3

Method

The participants were 361 (103 male and 257 female) undergraduate students at a large Midwestern U.S. university. Most participants were Caucasian (83.9%) and all participants received extra course credit in exchange for their participation.

The design, stimulus materials, measures, and procedures were identical to those in Priming Study 1, except that participants only completed one type of self-construal scale. This created a 2 (I-prime, we-prime) by 2 (independent or interdependent self-construal) independent groups design. The scale reliabilities were similar to those reported in Studies 1 and 2 (interdependence $\alpha = .72$ and independence $\alpha = .81$).

TABLE 2
The Effects of Situational Priming on Self-Construal

Self-construal type	*Independent prime*	*Interdependent prime*
Study 1		
Independent	4.29	4.19
Interdependent	3.09	3.13
Study 2		
Independent	4.17	4.13
Interdependent	3.18	3.05
Study 3		
Independent	4.16	4.14
Interdependent	3.09	3.03

NOTE: No statistically significant differences between priming conditions were observed.

Results

The data were analyzed with a 2-way ANOVA. Neither priming nor the anticipated prime by scale type interaction was significant ($F < 1.00$, $\eta^2 < .01$). Excluding non-Caucasians did not alter the results. The statistical power was .60 for small effects ($r = .10$) and greater than .995 for medium effects ($r = .24$). These results indicate, with 95% confidence, that $\eta^2 < .022$. The means for all three priming studies are presented in Table 2.

Discussion

Taken together, these data suggest that priming has little impact on self-report self-construal scale scores. Across the three studies, the weighted mean effect for priming on self-construals was $r = .019$, $p = ns$. The failure to obtain significant priming effects is not likely to be attributable to problems with the priming induction because identical procedures worked in previous priming experiments (Brewer & Gardner, 1996; Gardner et al., 1999; Trafimow et al., 1991; Trafimow et al., 1997; Ybarra & Trafimow, 1998) and because two different priming manipulations were conducted in Study 2. Low statistical power and item order effects can also be eliminated as explanations for the lack of findings in Study 3. The scale reliabilties were such that attenuation due to measurement error is not a likely explanation for the failure to find substantial effect. Although

point null hypotheses cannot be proven, the power analyses and confidence intervals provide evidence that the effects of the priming induction on self-construal scales are trivial. Instead, the most plausible explanation for the results is that self-construal scales, at least in Western cultures, measure stable, trait-like constructs. These findings have two important implications.

First, the results are informative regarding the construct validity of self-construal scales. Recall that the independent self is conceptualized as stable, whereas the interdependent self is situation specific and malleable. The results of the priming studies suggest that the scales measuring both types of self-construals reflect stable constructs impervious to situational priming. These findings are consistent with the construct validity of independent but not interdependent self-construal.

Further, research investigating self-construals with methods other than the self-construal scales finds that self-construals are consistently influenced by priming (e.g., Brewer & Gardner, 1996; Gardner et al., 1999; Trafimow et al., 1991; Trafimow et al., 1997) and situational variation (Kanagawa et al., 2000). The weighted mean effect for priming on self-construals in these studies is $r = .389$, $p < .001$. Comparing the results of research using alternative measures to the current results suggests that different measures of the same construct fail to produce parallel results. Thus, the results of the current priming studies, when compared to previous priming studies, imply that self-construal scales lack convergent validity.

A second implication of the priming studies is that, contrary to our earlier speculation, subtle situational priming cannot account for the cross-study inconsistencies documented in the meta-analysis. Therefore, an alternative explanation for the meta-analytic results is needed. We will next next question if the inconsistencies in previous results might be attributable to instability and measurement confounds in self-construal scales.

The Psychometric Properties of Self-Construal Scales

Self-construals may appear to lack construct validity because the scales used to measure self-construals may be problematic. Both the meta-analysis and the priming studies raise serious concerns over the construct validity of self-construal scales. Findings of heterogeneity in meta-analyses may signal problems with measurement validity (Fiske, 2002). Scholars have also raised concerns about the use of self-report scales to study culture (Fiske, 2002; Kanagawa et al., 2001; Kitayama, 2002; Markus & Kitayama, 1998). A careful examination of the literature reveals additional reasons for questioning the validity of self-construal scales.

First, it is curious that there are three distinct scales in common use (Gudykunst et al., 1996; Leung & Kim, 1997; Singelis, 1994). Each subsequent scale appears to be a refinement of earlier measures. For example,

TABLE 3
Self-Construal Items in the Leung and Kim (1997), Singelis (1994), and Gudykunst et al. (1996) Scales

	Scale author(s)		
Scale item	*L&K*	*S*	*G*
Independent self-construal items			
It is important for me to act as an independent person.	X		X
I should be judged on my own merit.	X		X
I voice my opinions in group discussions.	X		X
My personal identity, independent of others, is very important to me.	X	X	X
I prefer to be self-reliant rather than dependent on others.	X		X
I act as a unique person, separate from others.	X		X
I enjoy being unique and different from others	X	X	X
I don't like depending on others.	X		X
Being able to take care of myself is a primary concern for me.		X	X
I take responsibility for my own actions.	X		X
I don't change my opinions in conformity with those of the majority.	X		
Speaking up in a work/task group/class is not a problem for me.	X	X	
Understanding myself is a major goal in my life.	X		
I enjoy being admired for my unique qualities.	X		
I have an opinion about most things: I know what I like and I know what I don't like.	X		
If there is a conflict between my values and the groups of which I am a member, I follow my values.			X
I should decide my future on my own.			X
What happens to me is my own doing.			X
I assert my opposition when I disagree with members of my group.			X
I don't support a group decision when I know it is wrong.			X
Having a lively imagination is important to me.	X	X	
I'd rather say "no" directly than risk being misunderstood.		X	
I am comfortable being singled out for praise or rewards.		X	
I am the same person at home that I am at school.		X	

TABLE 3 Continued
Self-Construal Items in the Leung and Kim (1997), Singelis (1994), and Gudykunst et al. (1996) Scales

	Scale author(s)		
Scale item	*L&K*	*S*	*G*
I act the same way no matter who I am with.		X	
I feel comfortable using someone's first name soon after I meet them, even when they are much older than I am.		X	
I prefer to be direct and forthright when dealing with people I've just met.		X	
I value being in good health above everything.		X	
Interdependent Self-Construal Items			
I feel uncomfortable disagreeing with my group.	X		
I conceal my negative emotions so I won't cause unhappiness among the members of my group.	X		
I will stick with my group, even through difficulties.	X		
My relationships with those in my group are more important than my personal accomplishments.	X	X	X
My happiness depends on the happiness of those in my group.	X	X	X
I often consider how I can be helpful to specific others in my group.	X		
I am careful to maintain harmony in my group.	X	X	X
When with my group, I watch my words so I won't offend anyone.	X		
I would sacrifice my self-interests for the benefit of my group.	X	X	X
I will stay in a group if they need me, even if I'm not happy with the group.		X	X
I try to meet the demands of my group, even if it means controlling my own desires.	X		X
It is important to consult close friends and get their ideas before making decisions.	X		X
I consult with co-workers on work-related matters.			X
I consult with others before making important decisions.			X
I should take into consideration my parents' advice when making education and career plans.	X	X	
It is better to consult with others and get their opinions before doing anything.			X

TABLE 3 Continued
Self-Construal Items in the Leung and Kim (1997), Singelis (1994), and Gudykunst et al. (1996) Scales

	Scale author(s)		
Scale item	*L&K*	*S*	*G*
I respect decisions made by my group.		X	X
I respect the majority's wishes in the groups of which I am a member.			X
I act as fellow group members prefer I act.	X		X
The security of being an accepted member of a group is very important to me.	X		
I try to abide by customs and conventions at work.			X
If my brother or sisters fails, I feel responsible.	X	X	
I have respect for the authority figures with whom I interact.		X	
I would offer my seat on the bus to my professor.		X	
I respect people who are modest about themselves.		X	
Even when I strongly disagree with group members, I avoid an argument.		X	

NOTE: L&K marks Leung and Kim (1997) scale items, S denotes Singelis (1994), and G refers to Gudykunst et al. (1996). Similarly worded items are listed only once, and hence not all item wordings are exact.

Gudykunst et al. (1996) included Singelis (1994) items along with new items, and Leung (1997), in the original version of Leung and Kim (1997), included both Singelis and Gudykunst items. Neither Gudykunst et al., nor Leung provide an explanation for why scale revisions were necessary, but the evolution of the scales suggests that subsequent authors were dissatisfied with previous measures. A comparison of the items comprising these three scales is provided in Table 3.

Second, given the quantity of research using these scales, there has been surprisingly little validation work published, and the evidence that exists suggests reasons for concern. Of the three main scales, only the original validation of Singelis (1994) and an independent validation of the Gudykunst scale (Hackman et al., 1999) have been published. Both Singelis (1994) and Hackman et al. (1999) argue for the validity of the respective scales, but in both cases, their conclusions seem at odds with their data. Singelis found that a two-factor model fit better than a one-

factor model. Goodness of fit tests for the hypothesized two-factor model, however, yielded results that were less than ideal (goodness of fit index = . 853 and .809). Hackman et al. report a confirmatory factor analysis (CFA) of the Gudykunst scale and found that, in order to make the model fit, they had to disregard tests of parallelism completely, discard six items, and correlate several error terms in internal consistency. Published CFAs of the Leung and Kim scale have not reported tests of parallelism at all (e.g., Kim et al., 1998; Tasaki et al., 1999). Ohashi (2000), however, considered both the internal consistency and parallelism of the Leung and Kim scale, and reported dropping 10 independent (of 15) items and 11 (of 14) interdependent items in order to obtain acceptable fit. Additional evidence of problematic items is reported in Redford (1998) and Youn (2000). Redford dropped nine items from the Gudykunst scales because the items did not have at least .3 factor loadings in exploratory factor analysis (EFA), and Youn reported dropping six Singelis items because of low item-total correlations.

Most researchers using self-construal scales, however, only consider scale alpha reliability as evidence for scale adequacy. Alpha, however, can only be meaningfully interpreted when a scale is unidimensional and free from measurement confounds (Shevlin, Miles, Davies, & Walker, 2000). This is because confounds (reflected by correlated item error terms) can artificially inflate alpha.[3] Shevlin et al. (2000) note that EFA is not sufficient to document dimensionality. Instead CFA or structural equation modeling (SEM) is required. Because Gudykunst et al. (1996) and Leung (1997) relied exclusively on EFA, their results cannot be definitive. Further, because of correlated error terms, Hackman et al.'s (1999) data suggest the possibility of confounded measurement and inflated alphas.

Third, concerns over dimensionality of existing self-construal scales are exacerbated by findings that both independent and interdependent self-construals are multidimensional. Kashima et al. (1995) provided evidence for three types of self-construals (individualistic, collective, and relational) and further suggested that the individualistic (independent) dimension has two subdimensions (agency and assertiveness) that are orthogonal at the individual level. Consistent with Kashima et al., Cross, Bacon, and Morris (2000) provide convincing evidence for two distinct forms of interdependent self-construal (collectivism-based interdependence and relational interdependence). Sato and McCann (1998) found that items on the Singelis scale load on four different factors (sensitivity, autonomy, achievement, and attachment). Finally, Fiske (2002) contends that 4 types of interdependence and 10 types of independence may be conflated in self-construal scales. Taken together, these data suggest that self-construal scale items may measure more than two constructs.

Finally, examination of item content in Table 3 raises serious questions related to face validity and dimensionality. Some items appear to directly

address self-concept whereas others address typical behaviors and the theoretically distinct constructs of face, power distance, conformity, communicative directness, and communication apprehension. What construct other items may be tapping is unclear (e.g., "I value being in good health above everything," Singelis, 1994). Thus, a careful consideration of item content also suggests that self-construal items may be measuring more than two constructs.

MEASUREMENT STUDIES

As argued above, multiple reasons exist to question the validity of existing self-construal scales. If measures of independent and interdependent self-construals actually assess multiple constructs (i.e., the measures are confounded) or if the internal and cross-structures of the scales are highly unstable, or both, the scales would not be expected to yield consistent results explaining the high degree of variability in the results of the meta-analysis. To test this possibility, five measurement studies are reported.

Measurement Study 1

The dimensionality of self-construal scales was initially tested with the data from Priming Study 1 to see if measurement problems might exist. In that study, 121 participants completed the Leung and Kim (1997) scale. Responses to the scale were first analyzed using CFA with Hunter and Hamilton's (1987) *PACKAGE, Version 2.0*. This program (a) provides factor loadings based on a centroid solution, (b) computes predicted correlations based on factor loadings and the model specified, and (c) calculates deviations between predicted and obtained correlations. The magnitude of these deviations are tested against sampling error to determine fit. A two-factor solution with all independent self-construal items loading on one factor and all interdependence items loading on the other factor was specified a priori. Tests of both internal consistency and parallelism were conducted. A detailed description of this procedure may be found in Hunter and Gerbing (1982).

The test of internal consistency for independent self-construal yielded six significant deviations at $p < .05$ (4 at $p < .01$) distributed across 7 of 15 items. For interdependent self-construal, the test of internal consistency produced 11 significant deviations at $p < .05$ (5 at $p < .01$) distributed across 10 of 14 items. The parallelism test resulted in 31 significant deviations at $p < .05$ (13 at $p < .01$) distributed across 25 of the 29 items. Across tests, only one interdependence and two independence items failed to produce significant deviations for the predicted model. Together, these results indi-

cate a substantial departure from the a priori model and suggest that serious measurement problems may exist in the Leung and Kim (1997) scale.

To obtain estimates of global fit, the a priori two-factor model was retested using maximum likelihood CFA methods with EQS structural equations software. The fit indexes examined included the comparative fit index (CFI), the goodness of fit index (GFI), and the root mean squared error of approximation (RMSEA). CFI values range from 0 to 1.0, with values greater than .90 indicating close fit (Bentler, 1995). GFI is a measure of the relative amount of variances and covariances in a sample that are accounted for by the implied model (Jöreskog & Sörbom, 1984). A GFI of greater than .9 is conventionally considered to indicate an acceptable fit (Byrne & Cambell, 1999). With RMSEA, values less than .05 indicate close fit, values between .05 and .08 indicate reasonable fit, values between .08 and .10 indicate moderate fit, and values greater than .10 indicate unacceptable fit (Browne & Cudeck 1993).

Consistent with the results from PACKAGE, the fit of the two-factor model to the data with the EQS analysis was poor ($\chi^2_{376} = 726.22$, $p < .001$, CFI = .559, GFI = .687, RMSEA = .088). These findings are replicated and extended in Measurement Studies 2 and 5 below.

Measurement Study 2

The second measurement study replicates the results of the Leung and Kim scale (1997) with larger samples and cross-cultural data. These data test if the measurement problems identified in Measurement Study 1 will be observed in other data sets collected in the same and different locations. The research strategy in Measurement Study 2 is to assess the fit of the two-factor self-construal model within two different countries (the U.S. and Japan) with CFA using PACKAGE and EQS.

Method and Results

The data used in Study 2 were originally collected by Ohashi (2000). The participants included 223 students from Rikkyo University in Tokyo, Japan and 230 undergraduate students from the same large Midwestern U.S. university as in Study 1. All participants completed the Leung and Kim (1997) self-construal scale. The questionnaire was translated from English to Japanese and back translated from Japanese to English by two bilinguals. The Japanese version of the questionnaire was distributed to Japanese students and the English version of the questionnaire was distributed to participants from the United States. A complete description of the methods is presented in Ohashi (2000).

Responses to the Leung and Kim (1997) self-construal scale were sub-

mitted to confirmatory factor analysis with PACKAGE. In the Japanese data, the test of internal consistency for independent self-construal yielded 20 significant deviations at $p < .05$ (9 at $p < .01$) distributed across all but 1 of the 15 items. For interdependent self-construal, the test of internal consistency produced 18 significant deviations at $p < .05$ (3 at $p < .01$) distributed across 13 of 15 items. The parallelism test resulted in 50 significant deviations at $p < .05$ (28 at $p < .01$) distributed across all 30 items. For the U.S. data, the test of internal consistency for independent self-construal yielded 10 significant deviations at $p < .05$ (4 at $p < .01$) distributed across 11 of 15 items. For interdependent self-construal, the test of internal consistency produced 8 significant deviations at $p < .05$ (3 at $p < .01$) distributed across 11 of 15 items. The parallelism test resulted in 63 significant deviations at $p < .05$ (37 at $p < .01$) distributed across 28 of 29 items.

To obtain estimates of global fit, the a priori two-factor model was retested using maximum likelihood CFA methods with EQS. The fit of the two-factor model was poor in both the data from Japan ($\chi^2_{376} = 1050.64$, $p < .001$, CFI = .597, GFI = .733, RMSEA = .090) and the data collected in the U.S. ($\chi^2_{376} = 834.73$, $p < .001$, CFI = .697, GFI = .788, RMSEA = .073). Thus, the inconsistency of the Leung and Kim (1997) self-construal scale with the a priori two-factor model reported in Study 1 was replicated with cross-cultural data. Together, the first two measurement studies demonstrate that serious measurement problems exist with the Leung and Kim (1997) scale.

Measurement Study 3

The third measurement study assessed the fit of the Gudykunst et al. (1996) scale. The Gudykunst scale was tested to determine if the measurement problems observed in Studies 1 and 2 can be attributed to idiosyncracies in the Leung and Kim (1997) scale or whether the problems are more general. Because the Leung and Kim scale was a refinement of the Gudykunst scale, one might anticipate an even poorer fit in the Gudykunst scale.

The data for this study were originally collected by Lapinski (1995) and published as Lapinski and Levine (2000). The data were collected from 323 undergraduate students at the University of Hawaii. Participants completed a questionnaire containing, among other things, the 30-item Gudykunst et al. (1996) self-construal scale. Previous reports of the data included only item-total correlations, upon which one independent self-construal item was deleted. A complete description of the procedures can be found in Lapinski and Levine (2000).

Responses to the Gudykunst et al. (1996) scale were submitted to confirmatory factor analysis with PACKAGE. The test of internal consistency

for independent self-construal yielded 21 significant deviations at $p < .05$ (12 at $p < .01$) distributed across all but 1 of the 15 items. For interdependent self-construal, the test of internal consistency produced 22 significant deviations at $p < .05$ (11 at $p < .01$) distributed across 14 of 15 items. The parallelism test resulted in 50 significant deviations at $p < .05$ (26 at $p < .01$) distributed across 27 of the 30 items. Across tests, every item produced significant deviations from the predicted model.

Estimates of global fit were again obtained from a maximum likelihood CFA with EQS. The fit was clearly unacceptable ($\chi^2_{404} = 1955.47$, $p < .001$, CFI = .277, GFI = .656, RMSEA = .109). Together with the results of PACKAGE, these tests provide strong evidence of severe measurement problems in the Gudykunst scale. The data also suggest that measurement problems are not confined to the Leung and Kim (1997) scale. Although the results of Studies 1 and 2 suggest serious measurement problems with the Leung and Kim (1997) scale, the fit of the Gudykunst scale is even worse.

Measurement Study 4

The fourth measurement study assessed the fit of the Singelis (1994) scale. The Singelis scale was tested to determine if the measurement problems observed in the Leung and Kim (1997) scale and the Gudykunst et al. (1996) scale are also evident in the Singelis scale. Because both of the newer scales are refinements of the Singelis scale, one might anticipate an even poorer fit in the Singelis scale. However, because Singelis conducted CFAs and reported fit superior to the values reported above, the Singelis scale might be expected to fit the data better than its rivals.

Method and Results

Participants included 214 students from Tongmyong University of Information Technology located in Pusan, South Korea; 206 students from two universities (male students were from Hankuk University of Foreign Studies and all female students were from Ehwa Women's University) in Seoul, South Korea; 126 students from Seinan Gakuin University in Japan, and 204 undergraduate students from the same large Midwestern U.S. university as in Study 1. All participants completed the Singelis (1994) self-construal scale. The questionnaire was translated from English to both Korean and Japanese, and it was then back translated. The Korean version of the questionnaire was distributed to Korean students, the Japanese version was given to the Japanese participants, and the English version of the questionnaire was distributed to the U.S. participants.

Responses to the Singelis (1994) self-construal scale were submitted to confirmatory factor analysis with PACKAGE. In the Pusan data, the test

of internal consistency for independent self-construal yielded 6 significant deviations at $p < .05$ (2 at $p < .01$) distributed across 8 of the 12 items. For interdependent self-construal, the test of internal consistency produced 18 significant deviations at $p < .05$ (11 at $p < .01$) distributed across 11 of 12 items. The parallelism test resulted in 30 significant deviations at $p < .05$ (16 at $p < .01$) distributed across 23 of 24 items. In the Seoul data set, the test of internal consistency for independent self-construal yielded 11 significant deviations at $p < .05$ (4 at $p < .01$) distributed across 9 of the 12 items. For interdependent self-construal, the test of internal consistency was clean with no significant deviations observed. The parallelism test resulted in 30 significant deviations at $p < .05$ (8 at $p < .01$) distributed across 22 of 24 items. In the Japanese data, the test of internal consistency for independent self-construal yielded 14 significant deviations at $p < .05$ (7 at $p < .01$) distributed across 11 of the 12 items. For interdependent self-construal, the test of internal consistency produced 10 significant deviations at $p < .05$ (2 at $p < .01$) distributed across 11 of 12 items. The parallelism test resulted in 33 significant deviations at $p < .05$ (17 at $p < .01$) distributed across 22 of 24 items. For the U.S. data, the test of internal consistency for independent self-construal yielded 9 significant deviations at $p < .05$ (2 at $p < .01$) distributed across 9 of 12 items. For interdependent self-construal, the test of internal consistency produced 9 significant deviations at $p < .05$ (5 at $p < .01$) distributed across 11 of 12 items. The parallelism test resulted in 16 significant deviations at $p < .05$ (3 at $p < .01$) distributed across 15 of 24 items.

To obtain estimates of global fit, the a priori two-factor model was retested using maximum likelihood CFA methods with EQS. Consistent with the results from PACKAGE, the fit of the two-factor model to the data with the EQS analysis was poor in all four data sets: the Pusan data ($\chi^2_{251} = 607.46$, $p < .001$, CFI = .596, GFI = .786, RMSEA = .236), Seoul data ($\chi^2_{251} = 544.76$, $p < .001$, CFI = .440, GFI = .825, RMSEA = .076), the Japanese data ($\chi^2_{251} = 695.06$, $p < .001$, CFI = .251, GFI = .694, RMSEA = .268), and the data collected in the U.S. ($\chi^2_{251} = 624.24$, $p < .001$, CFI = .635, GFI = .798, RMSEA = .196).

Measurement Study 5

The final measurement study further examines the Leung and Kim (1997) scale's performance with cross-cultural data.[4] For self-construal scales to have utility in cross-cultural research, some degree of invariance across cultures is necessary (Byrne & Campbell, 1999; Church & Lonner, 1998; Ghorpade, Hattrup, & Lackritz, 1999; Paunonen & Ashton, 1998). "Invariance testing across groups, however, assumes well fitting single-group

models" (Byrne, Shavelson, & Muthen, 1989, p. 456). Therefore, assessing fit in different cultures is a logical first step in testing if self-construal scales are portable across cultures (Byrne et al., 1989; also see Ghorpade et al., 1999).

The research strategy in Study 5 is to assess the fit of the two-factor self-construal model within three cultures. If a reasonable fit is obtained within all three groups, multiple-group analysis can be used to assess invariance across cultures (Byrne & Campbell, 1999; Paunonen & Ashton, 1998). If fit is problematic, exploratory analyses will be conducted to assess if an alternative multiple-groups model is plausible.

Method

The data used in Study 5 were originally collected by Park (1998) and published in Park and Levine (1999). The participants included 148 students from Chung-Ang University located in Seoul, South Korea; 141 undergraduate students from the University of Hawaii; and 150 undergraduate students from the same large Midwestern U.S. university as in Study 1. All participants completed the Leung and Kim (1997) self-construal scale. The questionnaire was translated from English to Korean and back translated from Korean to English by two bilinguals. The Korean version of the questionnaire was distributed to the Korean students and the English version of the questionnaire was distributed to the participants from Hawaii and the Midwestern United States. A complete description of the methods is presented in Park and Levine (1999).

Results and Discussion

Separate CFAs were conducted for each culture with a maximum likelihood estimation on EQS. As in Study 1, the data from the Midwestern U.S. showed substantial departure from the model. All indices were well below the recommended levels, $\chi^2(375, N = 150) = 889.735$, $p < .001$, CFI = .647, RMSEA = .096, GFI = .647. Similarly, the data failed to fit the model in the Korean sample, and once again all indices were well below the recommended levels, $\chi^2(376, N = 148) = 1127.987$, $p < .001$, CFI = .603, RMSEA = .117, GFI = .651. For the Hawaii data, the model did not fit the data adequately; $\chi^2(375, N = 141) = 969.540$, $p < .001$, CFI = .589, RMSEA = .107, GFI = .672. The poor fit in the Hawaii data is especially troubling because the scale was developed in Hawaii. Thus, the lack of fit observed in Studies 1 and 2 was replicated, and the results indicate that the a priori two-factor model is inconsistent with the data in all three cultures tested.

From the above analyses, it is apparent that some modification in model specification is needed in order to determine a model that better represents the sample data. However, because previous research failed to show an adequate two-factor model (Hackman et al., 1999; Singelis, 1994; Stud-

TABLE 4
The Relative Fit of 2- to 8- factor Models with EFA in Measurement Study 3

	Number of factor specified						
Fit index	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Korean sample							
Chi-square	902	704	660	619	536	378	446
df	349	322	296	271	247	224	202
CVI	7.31	6.37	6.42	6.49	6.25	5.48	6.25
RMSEA	.104	.090	.092	.094	.089	.069	.091
Hawaiian sample							
Chi-square	845	792	554	628	430	386	415
df	349	322	296	271	247	224	202
CVI	7.27	7.27	5.92	6.83	5.76	5.77	6.30
RMSEA	.101	.102	.079	.097	.073	.072	.087
Midwestern U.S. sample							
Chi-square	709	770	543	481	373	305	268
df	349	322	296	271	247	224	202
CVI	5.91	6.69	5.51	5.43	5.03	4.88	4.93
RMSEA	.083	.097	.075	.072	.059	.049	.047

NOTE: All chi-square values are statistically significant at $p < .001$. Smaller values of CVI indicate a better fit. RMSEA values less than .05 indicate close fit, values between .05 and .08 indicate reasonable fit, values between .08 and .10 indicate marginal fit, and values greater than .10 indicate unacceptable fit.

ies 1 & 2 above) and is even suggestive of more than two factors (Cross et al., 2000; Kashima et al., 1995; Sato & McCann, 1998), the next step employed exploratory factor analyses (EFA). These analyses explore the possibility that there may be more than two factors in the self-construal scale and address the issues of stability in factor number and composition across cultures.

Maximum likelihood EFAs were conducted with FCAP (Browne & Cudeck, 1991) for each culture. Two-to eight-factor solutions with an oblique rotation (Direct Quartimin) were imposed and two types of fit indices were used to find the best fitting solution. The cross-validation index

(CVI) is intended to identify models that will perform optimally in future samples. Smaller values for CVI indicate better fit. The RMSEA was also considered.

Table 4 shows the fit indices for each factor model within each culture. In Korea, a seven-factor model showed a reasonable fit. Six- and seven-factor models showed reasonable fit in Hawaii. In the Midwestern U.S. sample, the six-factor model produced a reasonable fit and the seven-factor model showed close fit. In each case, six- or seven- factor models resulted in considerably better fit than the a priori two-factor model.

Although the number of factors suggested were similar across cultures, the items comprising these factors varied dramatically across samples. In Korea, interdependent items formed two factors, independent items formed four factors, and one factor was composed of both types of items. A somewhat similar pattern emerged in the Hawaii data where interdependent items again formed two factors, and independent items formed five factors. However, the groupings of specific items within factors varied considerably between Hawaii and Korea. For example, items comprising one factor in the Korean data split across three separate factors in the Hawaii data. In the Midwestern sample, interdependent items formed four factors, independent items formed two factors, and one factor contained both types of items. Across cultures, only six interdependent items and five independent items belong to the same factors across cultures, although they group with each other differently.[5] Examination of item content did not aid in making sense of the discrepant solutions.

Based on these data, together with the findings of previous research (e.g., Cross et al., 2000; Kashima et al., 1995; Sato & McCann, 1998) and the repeated failures of the two-factor model reported above, it is reasonable to conclude that self-construal scales form more than the proposed two dimensions. It is also reasonable to infer that self-construal scales do not have anything remotely close to an invariant factor structure across cultures. The data from different cultures show that items form different factors across cultures. Thus, these results provide strong evidence that the Leung and Kim (1997) self-construal scale is radically multidimensional, highly unstable, and ill-suited for use in cross-cultural research.

No effort was made to refine the scales to achieve model fit across cultures. This decision was based on a number of considerations. First, in previous research, such refinement required drastic methods. For example, Ohashi (2000) had to drop more than 75% of the items to obtain fit, and Hackman et al. (1999) dropped some items and loosened model constraints considerably. No less extreme procedures would be needed to obtain model fit here. As Byrne et al. (1989) noted, post hoc analyses with covariance structure models are problematic and tend to capitalize on chance. We therefore do not believe that such radical refinement is justified. Second, past refinements (e.g., Gudykunst et al., 1996; Leung & Kim, 1997)

have proven grossly inadequate, and we see little reason for optimism regarding the current prospects for refinement of the two-factor model. Finally, we subscribe to a theory-driven confirmatory approach to measurement-model testing that is incompatible with the type of post hoc model fitting that would be required.

The results of the measurement studies have important implications for the interpretation of the meta-analytic results reported above. First, drawing inferences about mean differences between cultures requires at least some degree of measurement invariance across cultures (Byrne & Campbell, 1999; Church & Lonner, 1998; Ghorpade et al., 1999; Paunonen & Ashton, 1998). Applied to self-construals, this means that in order to make meaningful cross-cultural comparisons of mean scores on self-construal scales, self-construal scales must provide a reasonable fit to the theoretically based measurement model within cultures, and the models must further be at least minimally constant across cultures. Because the measurement studies show that neither condition is met, valid cross-cultural comparisons involving self-construal scales are not possible.

Second, the highly unstable nature of self-construal scales may explain the radical across-study variation observed in the meta-analysis. The measurement studies reported here suggest that self-construal scales do not consistently measure the same constructs across data sets. This offers a reasonable explanation for why the meta-analysis showed that the effects observed in different studies came from different populations. Thus, it is plausible that the instability in self-construal findings stem from instability in popular self-construal scales.

GENERAL DISCUSSION

This article began by reporting a meta-analysis of published cross-cultural self-construal research that used popular self-report scales. The results allow for several conclusions. First, existing data do not support the conclusions that either self-construals reflect individual-level cultural orientation or that self-construals mediate and explain cross-cultural differences. Second, the results of previous research investigating cross-cultural differences in self-construals vary radically from study to study, and the literature can only be accurately described as strikingly inconsistent. Third, strong evidence exists that some unidentified moderators, confounds, or artifacts have contaminated previous findings.

Next, three priming studies were reported. The results indicated that situational priming does not explain the across-study variation observed in the meta-analysis. Across the three studies, no evidence was found to support the hypothesis that situational priming affects self-construals as

measured by popular scales. Instead, the data suggest that self-construal scales measure stable trait-like constructs. Thus, self-construal scales are insensitive to the flexible nature of the interdependent self-construal. Further, because alternative measures of self-construal (e.g., the 20 questions test) are sensitive to situational priming, self-construal scales appear to lack convergent validity. Thus, the results of the three priming studies raise additional concerns regarding the construct validity of self-construal scales.

Finally, the paper reported five measurement studies. In each, the adequacy of self-construal scales was assessed with confirmatory factor analyses. A total of 11 separate samples were analyzed. The data were collected in a variety of locations by a variety of different researchers using three different versions of self-construal scales. The data were also analyzed with two different software packages using different approaches to CFA. Yet, in every case the results were the same. The data departed radically from the theoretically specified two-factor model. Further, exploratory analyses provided evidence that self-construal scales are radically multidimensional and highly unstable. These findings may explain the highly inconsistent results observed in the meta-analysis. Together, these findings point to severe and perhaps fatal flaws in the self-construal scales.

The present data lend substantial credence to Matsumoto's (1999) concerns over self-construal research. The data also seem to corroborate the arguments of Kanagawa et al. (2001), Kitayama (2002) and Markus and Kitayama (1998) that self-report scales may be ill-suited to study interdependent self-construal. Subsequent cross-cultural research should avoid existing self-construal scales, and future research may wish to rethink if closed-format self-report scales are the best way to assess self-concept in non-Western countries (cf. Fiske, 2002).

The most important question raised by these results, however, is whether the observed problems are solely attributable to the use of flawed scales or if the findings reflect more fundamental problems with the self-construal construct itself. Unfortunately, the current data cannot provide an empirical answer to this question since it is impossible to completely separate measures from the constructs being measured. Nevertheless, some speculation is warranted.

Perhaps the central theoretical issue centers around the conceptual definition of self-construal. Is it useful to think of interdependent and independent self-construals as the two types of self-construal most central to culture? If so, how are these best conceptualized?

For example, is the autonomous versus related-to-others distinction in self-concept conceptually and empirically equivalent to the stable versus flexible distinction in self-concept? Can we equate the self-in-relation-to-others with a collectivistic self-concept? Who constitutes those

ambiguous "others" and who comprises the constantly referenced "my group" in self-construal items?

The definitions of interdependent and independent self-construal commonly provided in the literature confound the current elements. For example, the independent self-construal is usually defined as a stable-autonomous self whereas the interdependent self is a situationally-variable self that emphasizes connections with others. Recent research suggests that these issues may need to be untangled before self-construals can have solid conceptual footing (Fiske, 2002). Kanagawa et al.'s (2001) data, for example, suggested that cultural differences may be more evident in self-concept stability-flexibility than in the specific content of the self. Other research (e.g., Kashima et al., 1995; Wang et al., 2000) suggested that cultural differences are most readily seen along autonomy and agency dimensions, whereas a relational dimension reflects gender differences more than cultural differences. Cross et al. (2000) recently distinguished between two types of interdependent self-construal: collectivism-based interdependence (in which the group or situation dictates behavior) and relational interdependence (in which the self is defined in terms of relationships with close others). These recent developments suggest that the previous theoretical basis, which guided the development of self-construal scales, presents an overly simplistic view of the self in relation to culture. Thus, the failure of self-construal scales may reflect deeper conceptual ambiguities in the self-construal construct. These ambiguities will need to be resolved before researchers can decide how self-construals are best measured.

As noted earlier, self-construals seem to be widely accepted, and, with only a few exceptions, scholars have been reluctant to criticize self-construals even when their data have deviated from their theoretically based predictions. Because of this wide acceptance and because our data suggest severe validity problems with self-construals, some might attempt to discredit or discount the current findings and conclusions. We summarize our response to anticipated critics below.

Generally, we believe the conclusion that self-construal scales lack validity is, on social scientific standards, unusually clear-cut. First, there is a fundamental, but underappreciated, asymmetry in statistical and logical force between corroborating evidence and falsifying evidence. Evidence for falsification allows for much greater confidence than does evidence for confirmation (Meehl, 1986). Second, there is an unusually high degree of consistency between various evidential units. The results of previous research, the meta-analysis, the priming studies, and the measurement studies all point to serious flaws in self-construal research involving self-construal scales. Isolated anomalies might be reasonably ignored or rationalized, but the quantity, diversity, severity, and ubiquitousness of the incriminating evidence defies dismissal. Finally, the conclusion that self-

construal scales lack validity is, by far, the more parsimonious and logically consistent account of the existing data. If our conclusions are accepted, then the data presented here provide a coherent picture. Attempts to salvage self-construal scales, however, will almost certainly require a set of tenuous and complex post hoc arguments.

Two specific counterarguments are anticipated. First, some might argue (cf. Eaton & Louw, 2000) that the failure to find the predicted cultural differences in the meta-analysis may be an artifact of the pervasive use of students as participants in the studies comprising the meta-analysis. Such an argument would hold that if previous studies had used older adults, then surely the predicted cultural differences would have been observed. Although this is an empirical question, we agree that a theoretically sound argument could be made that predicted cultural differences should be more evident in older adults than in college students. Contrary to this reasoning, however, Oyserman et al. (2002), report that students and adults do not differ widely in individualism and collectivism. Further, it is difficult to see how the use of college students could create the range of problems observed in the priming and measurement study results. It is much more plausible, we think, that the problems observed in the meta-analysis are, at least in part, attributable to invalid scaling.

A second argument, based on Cronbach and Meehl's (1955) nomological network approach to construct validity, might hold that previous self-construal research does in fact provide substantial evidence consistent with the construct validity of self-construal scales. This argument would note that theory specifies that interdependent and independent self-construals should correlate in a specified manner with a variety of outcome measures like embarrassment, conflict strategies, conversational constraints, motivation to comply, and so forth. Research investigating these links provides overwhelming evidence that self-construals do correlate with these outcome measures as predicted (e.g., Kim et al., 1994; Kim et al., 1996; Oetzel, 1998a, 1998b, 1999; Park & Levine, 1999; Park et al., 1998; Sharkey & Singelis, 1995; Singelis et al., 1999; Singelis & Brown, 1995; Singelis & Sharkey, 1995). The argument would continue that not only do these studies provide an impressive quantity of evidence consistent with the construct validity of self-construal scales but that such findings would be highly unlikely if self-construal scales were invalid. Thus, our conclusion that self-construal scales lack validity must be flawed.

Our reply to such an argument is multifaceted. First, the Cronbach and Meehl (1955) approach to construct validation can be criticized on the logical grounds that it is circular (Nunnally & Bernstein, 1994). Second, for a variety of reasons, correlations in the social sciences can have substantial spurious components and do not provide as strong of evidence as many social scientists think they do (Meehl, 1986). Third, the

vast majority of evidence used to support such an argument relies on correlations with other self-report scales. Hence, this evidence might merely reflect shared method variance. Consistent with this, recent authors have observed that self-construal scales seem to predict other self-reports with more consistency than alternative types of measures (Levine et al., 1999; Park, 2001; for an exception see Ellis & Wittenbaum, 2000). Thus, a Campbell and Fiske (1959) approach to construct validation would constitute a stronger test of construct validity. Finally, and most importantly, the apparent evidence for construct validity is likely an artifact of the radically multidimensional nature of the self-construal scales. Consulting Table 3 again, the items comprising self-construals seem to test a variety of constructs besides self-concept including altruism, communication directness, communication apprehension, conformity, and sensitivity to face concerns. So, for example, the finding that scores on interdependent self-construal are positively correlated with other-oriented conversational constraints (Kim et al., 1996) may indicate nothing more than the fact that other-oriented conversational constraints items were included in the interdependent self-construal scale.

CONCLUSIONS

The data presented here demonstrate the serious and persistent flaws in existing self-construal scales. The data show that self-construal scales do not reliably reflect the intended cultural differences. The results of self-construal research vary widely across studies, indicating the presence of powerful artifacts. The interdependent self-construal scale functions as a stable trait that fails to reflect situationally variable aspects of self-concept. Results obtained using self-construal scales fail to converge with the results obtained using alternative measures of the same construct. The intended two-factor measurement fails to fit the data, fails badly, and fails consistently. Finally, self-construal scales appear radically multidimensional and highly unstable. Whether these problems stem only from extremely bad measurement or whether they reflect more fundamental theoretical inadequacies is debatable. Nevertheless, the data presented here clearly demonstrate that the results obtained from self-construal scales are incompatible with the constructs the scales were designed to measure.

These results provide yet another example of two reoccurring lessons in social scientific research. First, it is always wise to validate measures prior to embarking on sustained programs of research. When commonly used measures are not validated, there is always the risk of whole literatures becoming invalidated and ending in costly dead ends. Second, con-

firmation biases can prove costly. Overlooking or rationalizing a theory's failures is usually, in the long run, counterproductive. In the words of Imre Lakatos, "Blind commitment to a theory is not an intellectual virtue: it is an intellectual crime" (1978, p. 1).

NOTES

1. This hypothesis is, of course, a null hypothesis. We recognize that traditional significance testing cannot prove a point prediction of no difference. However, confidence intervals and power analyses can provide evidence that an effect is so small as to be theoretically or practically meaningless (Cohen, 1990).

2. We wish to express our appreciation to Wendi Gardner for providing her instructions and stimulus materials, as well as for her helpful advice on conducting priming studies. We used the same instructions, format, font type, and so forth as Gardner et al. (1999).

3. Confounds can artificially inflate alpha if the confound is positively correlated with the construct being measured. If any two items are a function of two or more constructs that are positively related to each other, then the interitem correlations will be inflated by the spurious effects attributable to the confounding construct. These inflated interitem correlations will artificially inflate alpha and increase the probability that EFA will under-factor the matrix, especially when orthogonal rotations are used. Such confounds can be detected with CFA or SEM by the presence of positively correlated error terms or by the presence of significant deviations in residuals. Both indications of confounded measurement are present in the self-construal literature. Thus, the reliance on EFA with orthogonal rotations and alpha is highly problematic, and the results of these analyses are likely misleading.

4. We choose to examine the cross-cultural application of the Leung and Kim scale in particular (as opposed to the earlier scales) because this scale represents the most recent refinement of self-construal scales. Also, comparing the results of Measurement Studies 1 and 2 with Measurement Studies 3 and 4, the Leung and Kim (1997) scale seemed to provide a relatively better fit to the data than the Gudykunst et. al. (1996) or Singelis (1994) scales. Therefore, we reasoned that of the three scales, the Leung and Kim scale had the best chance of providing at least a minimal degree of invariance. Again, we made the methodological choice designed to give self-construals the best chance of succeeding.

5. Two sets of independent items consistently clustered together across cultures. The items "I don't like depending on others," "I take responsibility for my own actions," and "It is important for me to act as an independent person," always loaded on the same factor. Similarly, the items "I act as a unique person, separate from others" and "I enjoy being unique and different from others," consistently loaded on the same factor. However, these two sets of items were on different factors in the Korean and Hawaiian data, whereas all five items were on the same factor in the Midwestern sample. For independent self-construal items, three sets of two items consistently clustered together. The items "I feel uncomfortable disagreeing with my group" and "I conceal my negative emotions so I won't cause unhappiness among members of my group" formed one set. "I would sacrifice my self-interest for the benefit of my group" and "I try to meet the demands of my group, even if it means controlling my own desires" formed a second set. The third set consisted of "the security of being an accepted member of a group is very important to me" and "If my brother or sister fails, I feel responsible." Sets 2 and 3 formed one factor in the Korean data, distinct from the first set. Sets 1 and 3 clustered together in the Hawaii data, separate from set 2. Finally, the three sets loaded on three different factors in the Midwestern sample. The reader should note that four of the five sets showing consistent grouping were also adjacent scale items, raising the possibility that these apparent consistencies reflect response sets.

REFERENCES

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Brewer, M. B., & Gardner, W. (1996). Who is this "we"? Levels of collective identity and self-representations. *Journal of Personality and Social Psychology*, *71*, 83–93.

Brockner, J., & Chen, Y. R. (1996). The moderating roles of self-esteem and self-construal in reaction to the self: Evidence from the People's Republic of China and the United States. *Journal of Personality and Social Psychology*, *71*, 603–615.

Brockner, J., Chen, Y. R., Mannix, E. A., Leung, K., & Skarlcki, D. P. (2000). Culture and procedural fairness: When the effects of what you do depend on how you do it. *Administrative Science Quarterly*, *45*, 138–159.

Browne, M. W., & Cudeck, R. (1991). *FCAP: Factor analysis estimators with common asymptotic properties*. Minneapolis: University of Minnesota.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Thousand Oaks, CA: Sage.

Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, *30*, 555–574.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for equivalence of factor covariance mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.

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

Church, A. T., & Lonner, W. J. (1998). The cross-cultural perspective in the study of personality. *Journal of Cross-Cultural Psychology*, *29*, 32–62.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304–1312.

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

Cross, S. E. (1995). Self-construals, coping, and stress in cross-cultural adaptation. *Journal of Cross-Cultural Psychology*, *26*, 673–697.

Cross, S. E., Bacon, P. L., & Morris, M. L. (2000). The relational-interdependent self-construal and relationships. *Journal of Personality and Social Psychology*, *78*, 791–808.

Eaton, L., & Louw, J. (2000). Culture and self in South Africa: Individualism-collectivism predictions. *The Journal of Social Psychology*, *140*, 210–217.

Ellis, J. B., & Wittenbaum, G. M. (2000). The nature of verbal self-promotion depends on self-construal. *Communication Research*, *27*, 704–722.

Fiske, A. P. (2002). Using individualism and collectivism to compare cultures—A critique of the validity 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.

Ghorpade, J., Hattrup, K., & Lackritz, J. R. (1999). The use of personality measures in cross-cultural research: A test of three personality scales across two countries. *Journal of Applied Psychology*, *84*, 670–679.

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.

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.

Hunter, J. E., & Gerbing, D. W. (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.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI user's guide* (3rd ed.). Mooresville, IN: Scientific Software.

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

Kashima, Y., Yamaguchi, S., Kim, U., Choi, S.-C., Gelfand, M. J., & 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. (1994). Cross-cultural comparisons of the perceived importance of conversational constraints. *Human Communication Research*, *21*, 128–151.

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

Kim, M. S., Klingle, R. S., Sharkey, W. F., Park, H. S., Smith, D. H., & Cai, D. (2000). A test of a cultural model of patients' motivations for verbal communication in patient-doctor interactions. *Communication Monographs*, *67*, 262–283.

Kim, M. S., Sharkey, W. F., & Singelis, T. M. (1994). The relationship between individuals' 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). Cultural influences on the preferred forms of requesting and re-requesting. *Communication Monographs*, *65*, 47–66.

Kim, M. S., Smith, D. H., & Yueguo, G. (1999). Medical decision making and Chinese patients' self-construals. *Health Communication*, *11*, 249–260.

Kitayama, S. (2002). Culture and basic psychological processes—Toward a system view of culture: Comment on Oyserman et al. (2002). *Psychological Bulletin*, *128*, 89–96.

Kleinknecht, R. A., Dinnel, D. L., Kleinknecht, E. E., Hiruma, N., & Harada, N. (1997). Cultural factors in social anxiety: A comparison of social phobia symptoms and taijin kyofusho. *Journal of Anxiety Disorders*, *11*, 157–177.

Krull, D. S., Loy, M. H. M., Line, J., Wang, C. F., Chen, S., & Zhao, X. (1999). The fundamental fundamental attribution error: Correspondence bias in individualist and collectivistic cultures. *Personality and Social Psychology Bulletin*, *25*, 1208–1219.

Kuhnen, U., & Hannover, B. (2000). Assimilation and contrast in social comparisons as a consequence of self-construal activation. *European Journal of Social Psychology*, *30*, 799–811.

Lakatos, I. (1978). *The methodology of scientific research programmes*. Cambridge, UK: Cambridge University Press.

Lapinski, M. K. (1995). *Deception and the self: A cultural examination of information manipulation theory*. Unpublished master's thesis, University of Hawaii, Honolulu.

Lapinski, M. K., & Levine, T. R. (2000). Culture and information manipulation theory: The effects of self construal and locus of benefit on information manipulation. *Communication Studies*, *51*, 55–73.

Lee, A. Y., Aaker, J. L., & Gardner, W. L. (2000). The pleasures and pains of distinct self-construals: The role of interdependence in regulatory focus. *Journal of Personality and Social Psychology*, *78*, 1122–1134.

Leung, T. (1997). *Self-construals and situational dependency in conflict style preference*. Unpublished master's thesis, University of Hawaii, Honolulu.

Leung, T., & Kim, M. S. (1997). *A revised self-construal scale.* Honolulu: University of Hawaii at Manoa.

Levine, T. R., Lapinski, M. K., Banas, J., Wong, N., Hu, A. D. S., Endo, K., et al. (1999, November). *Self-construal and self-other benefit as determinants of deceptive message generation*.

Paper presented at the annual meeting of the National Communication Association, Seattle, WA.
Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253.
Markus, H. R., & Kitayama, S. (1998). The cultural psychology of personality. *Journal of Cross-Cultural Psychology*, *29*, 63–87.
Matsumoto, D. (1999). Culture and self: An empirical assessment of Markus and Kitayama's theory of independent and interdependent self-construals. *Asian Journal of Social Psychology*, *2*, 289–310.
Meehl, P. E. (1986). What social scientists don't understand. In D. W. Fiske & R. A. Shweder (Eds.), *Metatheory in social science* (pp. 315–338). Chicago: University of Chicago Press.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
Oetzel, J. G. (1998a). Culturally homogeneous and heterogeneous groups: Explaining communication processes through individualism and collectivism and self-construal. *International Journal of Intercultural Relations*, *22*, 135–161.
Oetzel, J. G. (1998b). Explaining individual communication processes in homogeneous and heterogeneous groups through individualism-collectivism and self-construal. *Human Communication Research*, *25*, 202–224.
Oetzel, J. G. (1998c). The effects of self-construal and ethnicity on self-reported conflict styles. *Communication Reports*, *11*, 133–144.
Oetzel, J. G. (1999). The influence of situational features on perceived conflict styles and self-construals in work groups. *International Journal of Intercultural Relations*, *23*, 679–695.
Ohashi, R. (2000). *High/low-context communication: Conceptualization and scale development*. Unpublished doctoral dissertation, Michigan State University, East Lansing.
Okazaki, S. (2000). Asian American and white American differences on affective distress symptoms: Do symptom reports differ across reporting methods? *Journal of Cross-Cultural Psychology*, *31*, 603–625.
Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*, 3–72.
Park, H. S. (1998). *Culture, self-construals, and the attitudinal and normative determinants of behavioral intention: A cross-cultural test of the theory of reasoned action*. Unpublished master's thesis, University of Hawaii at Manoa, Honolulu.
Park, H. S. (2001). Self-construals as motivating factors 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.
Park, H. S., Levine, T. R., & Sharky, W. F. (1998). Predicting behavioral intentions to recycle in Hawai'i: The theory of reasoned action and self-construals. *Communication Studies*, *49*, 196–208.
Paunonen, S. V., & Ashton, M. C. (1998). The structured assessment of personality across cultures. *Journal of Cross-Cultural Psychology*, *29*, 150–170.
Redford, P. (1998). Self-construal and anger action tendencies in Hong Kong and the United Kingdom. In W. J. Lonner, D. L. Dinnel, D. K. Forgays, & S. A. Hayes (Eds.) *Merging past, present, and future in cross-cultural psychology* (pp. 367–386). Lisse, The Netherlands: Swets & Zeitlinger.
Sato, T., & Cameron, J. E. (1999). The relationship between collective self-esteem and self-construal in Japan and Canada. *The Journal of Social Psychology*, *139*, 426–435.
Sato, T., & McCann, D. (1988). Individual differences in relatedness and individuality: An exploration of two constructs. *Personality and Individual Differences*, *24*, 847–859.
Sharkey, W. F., & Singelis, T. M. (1995). Embarrassabiity and self-construal: A theoretical integration. *Personality and Individual Differences*, *19*, 919–926.

Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences*, *28*, 229–237.

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

Singelis, T. M., Bond, M. H., Sharkey, W. F., & Lai, C. S. Y. (1999). Unpacking culture's influence on self-esteem and embarrassability: The role of self-construals. *Journal of Cross-Cultural Psychology*, *30*, 315–341.

Singelis, T. M., & Brown, W. J. (1995). Culture, self, and collectivist communication: Linking culture to individual behavior. *Human Communication Research*, *21*, 354–389.

Singelis, T. M., & Sharkey, W. F. (1995). Culture, self-construal, and embarrassability. *Journal of Cross-Cultural Psychology*, *26*, 622–644.

Tasaki, K. T., Kim, M. S., & Miller, M. D. (1999). The effects of social status on cognitive elaboration and post-message attitude: Focusing on self-construals. *Communication Quarterly*, *47*, 196–212.

Trafimow, D., Silverman, E. S., Fan, R. M., & Law, J. S. F. (1997). The effects of language and priming on the relative accessibility of the private self and the collective self. *Journal of Cross-Cultural Psychology*, *28*, 107–123.

Trafimow, D., Triandis, H. C., & Goto, S. G. (1991). Some tests of the distinction between private self and the collective self. *Journal of Personality and Social Psychology*, *60*, 649–655.

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

Wang, C. L., Bristol, T., Mowen, J. C., & Chakraborty, G. (2000). Alternative models of self-construal: Dimensions of connectedness-separateness and advertising appeals to the culture and gender-specific self. *Journal of Consumer Psychology*, *9*, 107–115.

Ybarra, O., & Trafimow, D. (1998). How priming the private self or collective self affects the relative weights of attitudes and subjective norms. *Personality and Social Psychology Bulletin*, *24*, 362–370.

Youn, I. (2000). The culture specificity of epistemological beliefs about learning. *Asian Journal of Social Psychology*, *3*, 87–105.