Validity Problems Comparing Values Across Cultures and Possible Solutions

Kaiping Peng and Richard E. Nisbett
University of Michigan—Ann Arbor

Nancy Y. C. Wong
University of Hawaii at Manoa

The authors argue that commonly used ranking and rating methods of value surveys may have low validity in cross-cultural value comparisons because participants' reports about values can be affected by factors such as cultural differences in the meaning of particular value terms as well as the possibility that some value judgments are based on social comparison or deprivation rather than on any "direct reading" of personal preferences. Four different value survey methods—ranking, rating, attitude scale, and behavioral scenario methods—were compared. It was found that ranking and rating methods of assessing differences between Chinese and Americans had low convergence with each other and with the criterion of cultural experts' independent judgments. Attitude questions had slight and nonsignificant convergence with the expert judgment criterion. A scenario method of value assessment, however, yielded reasonable criterion validity.

Cross-cultural value comparison has long been the main tool for understanding similarities and differences among people with different cultural and ethnic backgrounds (Allport, Vernon, & Lindzey, 1951; Hofstede, 1980; Kluckhohn, 1951; Morris, 1956; Rokeach, 1973, 1979; Williams, 1968). One of the most common practices in this area of research is to ask citizens of a nation or members of a culture to rate the importance of a standard set of values as "guiding principles in their lives" (e.g., Schwartz, 1992; Schwartz & Blisky, 1987, 1990), or to rank the rela-

tive "importance of" or "preference" for these values (e.g., Rokeach, 1973, 1979). The cross-cultural comparisons are then made based on the aggregated responses on each value or on some sets of values (e.g., "dimensions" or "hierarchies"). But there is a question as to how valid this approach is. Do the value differences based on aggregated self-reported individual responses really reflect differences in behavior and preference? Can we use value survey methodology as we please for cross-cultural comparison without considering possible limitations? Perhaps more important, what are the possible limitations of cross-cultural value comparison and what are the possible solutions?

Factors That Might Affect Criterion Validity of Cross-Cultural Value Comparisons

There are many factors that can affect validity of cross-cultural value comparisons. Some of them are obvious such as problems with the contents of value surveys and the representativeness of survey samples, but others that are less obvious are caused by individual cognitive process, generated by the survey methodology we are using.

Definitions of Values

Current definitions of value, we believe, may contribute to the validity problems of comparing values across cultures because they have two opposite impli-

cations in terms of value–behavior consistency. One type of definition implies that value should not necessarily be consistent with the behaviors, such as Kluckhohn's (1951) description of value as a "conception of the desirable" (p. 395), Williams's (1968) definition of value as "criteria or standards of preference," (p. 283), and Rokeach's (1973) statement of value as "an enduring belief that a specific mode of conduct or end-state is personally preferable to its opposite" (p. 5). It is our opinion that if a value is something that one desires or prefers, then it could be something that one does not have yet; by the same token, if value is only a standard or criterion, then it could mean that one has not successfully accomplished it.

However, the other type of value definition suggests a strong causal relation between a value and behavior. According to social adaptation theory (Kahle, 1984), values are types of social cognition and motivation that function to facilitate adaptation to one's environment. If values are "functions" for "adaptation," then they should have consequential effects on the value-related behaviors. In fact, this type of approach has explicitly claimed that the causal influence of value should flow from abstract values to mid-range attitudes to specific behaviors (Homer & Kahle, 1988).

We believe that there are shared beliefs about what are preferred modes of conduct and end-states within cultures, which could best be characterized by the concept of value. However, we do want to make a distinction between value as collective representations and values as subjective judgments made by individuals. A lesson of social cognition is that self-reports about attitude or value are subjective products of mental constructive processes, and that the results of these mental processes are subject to various heuristics, biases, and errors (see Sudman, Bradburn, & Schwarz, 1996; for an extensive review on this topic). This article is about the potential sources of error in assessing values across cultures by means of self-reports by individuals and possible ways to reduce these sources of error.

Cultural Differences in Constructing Meaning

When participants are asked to judge a set of abstract broad values as a guiding principle in their lives, their conceptions of these values could be idiosyncratic and could be very different among members of different cultures just as they are among individuals

within the same culture (Kahneman & Miller, 1986). Consider, for example, the value of equality. An African American might interpret it in terms of racial equality, a woman might think about gender equality, a working class or blue-collar worker might want to define it as economic equality, and a Chinese person might emphasize social equality. Hence, the discussion of value preferences or importance at highly abstract and general levels could be misleading because of tacit definitions that differ across cultures.

Social Comparison Processes in Judgments About Values

Evaluation of one's own beliefs or values is likely to be influenced at least in part by social comparison processes. The basic assumption of the value survey is that participants assess their own values according to their own beliefs, so that we could aggregate each participant's independent judgments and use the mean to represent the culture that the participant comes from. The problem is that people often make judgments about their own values in relation to their beliefs about other people who are salient to them, hence their responses are relative ones (Festinger, 1954). For example, an American woman might think she values the concept of respecting elders very much, because in comparison to her acquaintances she believes that she cares a lot about this value. Thus, her rating or ranking of this value might be higher than that of a Chinese woman, who, in comparison to her own people, doubts whether she values that concept as much as others do. However, by looking at other behavioral indicators, the Chinese person might be seen as much more respectful to elders than the American. For instance, the Chinese person might walk behind the elders, sit after the elders have done so, and never raise her head too high in front of the elders. Thus, although the two cultural groups in comparison might actually differ in the importance of certain values in their life, the relativity of social comparison process can reduce or even reverse the differences between them.

Deprivation-Based Preferences

People often express stronger preferences for something they lack, or believe themselves deficient in, than they do for things they have. For example, an American might think the value of respecting elders is very important because of a belief that elders have not been given sufficient respect in this culture; on the

other hand, a Chinese person might think the value of respecting elders is less important because the value has been taken for granted. Moreover, Chinese are more likely than Americans to have experienced that care of elders can be burdensome (Peng & Chen, 1989). Or consider the value of freedom, an American might take the state of freedom for granted, as a base from which other, more salient values could be compared. The citizen of a totalitarian country, however, might place an extremely high value on freedom because it is something highly desirable that is not possessed. Because the perception of deprivation is also relative (Crosby, 1976; Williams, 1975), it might produce quite misleading expressions of value preference.

Criteria of Cross-Cultural Value Comparison Validity

According to the American Psychological Association's (1974) *Standards for Educational and Psychological Tests,* three principal categories of validity are usually measured by different procedures, serve different purposes, and allow different inferences to be made from the measurement results (Anastasi, 1982; Cronbach, 1984; Peng, 1989). Criterion validity, an assessment of the degree to which interpretations, descriptions, and predictions on the basis of test scores or other measures can be supported by behavioral or empirical evidence such as participants' performances, diagnosis of behaviors, or experts' judgments (Messick, 1989), would seem to be by far the most relevant for determining whether value surveys are really informative with respect to cultural differences.

How would we know that value survey results have good criterion validity? For cultural differences as for individual differences within a culture, the best evidence that a verbal measure is valid might seem to be that it actually predicts people's behavior. If we can accurately describe and predict how people might behave in a specific situation on the basis of the scores of the value survey, then we can claim that these results do have criterion validity. In the case of cross-cultural value comparison, if the test scores suggest that one group of participants endorse a certain value significantly more than another group, and these two groups of people do indeed behave differently on average in a situation that plausibly reflects the given value, then we could conclude that the results have good criterion validity. However, there are many

theoretical and practical difficulties in achieving this ideal criterion validity.

These difficulties have been demonstrated at the level of efforts to assess individual differences. People may wish to behave in a particular way but be prevented from doing so for pragmatic reasons or because of felt normative pressure. It is a staple of the attitude–behavior literature that behavior is often linked only very weakly to attitudes. Even someone for whom health is a paramount value, for example, cannot perform all sorts of exercise, eat all sorts of health food, and so on. Any given behavioral measure of the health value is therefore likely to be a weak indicator of people's actual position on the dimension. Similarly, people may attend church in a given community not because of religious feeling but because of normative pressures to do so. There might therefore be little behavioral variation but substantial variation in actual values.

More generally, social psychological studies have demonstrated that a person's behavior in a certain situation may not reflect the person's values, attitudes, or beliefs very much but rather may be a function of the details of the situation. Whether a participant helps a confederate may be a function of the degree of hurry the participant is experiencing; whether a participant expresses some belief may have less to do with whether the participant holds the belief than with the degree of pressure placed on the participant to express it. Social psychological experiments sometimes can be difficult to replicate even within the same culture because situational factors are difficult to reproduce with sufficient subtlety from a reading of procedure sections. For example, in the early years of research on cognitive dissonance, some experimenters failed to get dissonance effects in high choice conditions because they failed to provide just the right mix of verbally stated assurances that the decision was up to the participant while producing substantial social pressure with voice tone and body language. All of these problems can be made worse in cross-cultural comparisons by the fact that small details may affect the constraints in very different ways in different cultures. Thus a failure to show that differences in reported values are reflected in comparable differences in measured behavior could be misleading. The measures of value might have validity that cannot be captured by behavioral tests because the behavioral tests are subtly different in the two cultures.

A less perfect but relatively easy way of determining a measurement's criterion validity is to compare

its results with those of other tests that are presumed to measure the same concepts or traits, a technique called convergent validity (Campbell & Fiske, 1959). When convergence is high, we could hope that this is because both sets of results have good criterion validity. The problem is that if the convergence is low, we do not know which test is invalid or whether both are invalid. Furthermore, when the convergence is assessed between two sets of the same or highly similar measurements to the same group of participants, the results should be considered more of a reliability assessment than a validity assessment. Only the convergence of different measurements for the same characteristics could be considered as an indication of validity (Campbell & Fiske, 1959). Investigators studying value differences have seldom compared each other's results even for the purpose of measuring convergence.

Another way of determining the criterion validity of cross-cultural value measures would be to see whether they correlate with a set of independent judgments made by cultural experts who specialize in the understanding of the cultures in question and who have had considerable systematic contact with people from the cultures being compared. This method seems not to have been used in value studies despite its obvious promise and relative simplicity as compared with arduous and possibly misleading behavioral measures. The possibilities for establishing a broad "nomological net" (Cronbach & Meehl, 1955) would seem to be greater for studies of cultural differences than for studies of individual differences. For many cultures, there is a range of historical, ethnographic, and literary material that may be drawn on in establishing validity criteria. If survey results are in accord with the generalizations prevalent in cultural studies, this provides genuine evidence that survey results are capturing something real about cultures and cultural differences.

Evidence of Low Criterion Validity of Cross-Cultural Value Comparison

In the field of cultural psychology, there is increasing evidence that what people say is important to them does not necessarily reflect their actual behaviors or preferences. For example, Stevenson and his colleagues (Stevenson, Lee, & Graham, 1993) surveyed three groups of more than 3,000 high school students from Minneapolis, Taipei, Taiwan, and Shendai, Japan. They asked their participants questions about how important mathematics, sports, and other activities were in their lives. They found that Asian children ranked sports as more important than American children did and ranked math as less important than their American counterparts. But the actual time spent on the two activities (playing sports, studying math), as well as their final math test scores, suggested that the opposite was true (Fuligni & Stevenson, 1994; Stevenson et al., 1993). In other research, it was found that Japanese students are more likely than Americans and Canadians to agree that people should keep quiet at the dinner table, not fight on the bus, not spit on the ground, and so on. But real life behavioral observations suggest the opposite behavioral differences (Iwata, 1991).

This inconsistency between value reports and behavior can also occur among subcultures within a nation. For example, value surveys comparing different ethnic groups in the United States have shown that some minority parents and children gave higher ratings than their European and Asian American counterparts to the importance of education. But the data indicating the length of time that individuals stay in school and time spent studying seem to indicate otherwise. This discrepancy has been explained in terms of real life or environmental constraints, and in terms of educational aspirations for some minority groups (Bratz & Levine, 1978; Hare & Castanell, 1985; Stevenson et al., 1990), but whatever the reason for the discrepancies, they pose a validity problem for value-comparison studies.

In this study, we first checked the convergence of value survey results within a single culture, namely the Chinese culture. The convergence between two sets of results of the same measurement for the same group of the participants is the test-retest reliability assessment, but the convergence between two test results of the same measurement for different groups of participants may be considered an indication of validity (albeit a weak indication).

Relatively high reliability of American value survey results has been reported by several authors. Data reported by Rokeach (1973) for American samples show the test-retest reliability of 18 terminal values (or idealized end-states of existence), range from .51 to .88 and for 18 instrumental values (or idealized codes of conduct), range from .45 to .70. There are also numerous reports showing that the rankings of values are significantly related to individual differences in many social attitudes and behaviors (e.g., Homer & Kahle, 1988; Pitts & Woodside, 1984;

Rokeach, 1973; Schwartz & Inbar-Saban, 1988; Tetlock, 1986; Toler, 1975).

However, we found that the value survey results for Chinese participants using Rokeach's Value Survey (Rokeach, 1973) have had relatively lower convergence. Although most of the studies investigating values of Chinese people have used author-modified Chinese versions of Rokeach's Value Survey (Appleton, 1970, 1979; see a review by Yang, 1986) or focused on more theoretical discussion of cultural differences (Bond, 1988; Ng et al., 1982), we have found four studies (five Chinese samples) published in the *International Journal of Psychology* that have used the full scale of Rokeach's Values Survey, hence providing a complete picture of Chinese rankings of Rokeach's values (Feather, 1986; Katz, Juni, & Shope, 1993; Lau, 1988; Lau & Wong, 1992). We calculated Spearman's rank correlation coefficients between each pair of samples, and these are presented in Table 1. The correlations among these Chinese samples ranges from low to moderate, with an average around .45 for Rokeach's terminal values and .52 for the instrumental values. It is very clear from Table 1 that the correlation coefficients of value rankings between one study and another, between one sample and another sample, one author and another author, even for the same author at different times for slightly different participant groups, reflect discouragingly mediocre convergence. The samples in these studies can scarcely be defined as anything other than samples of convenience, but it should be noted that the samples

are all of college students, which would lead us to expect higher correlations between samples than we would if samples were more diverse.

Even more ominous for hopes of validity are correlations presented in Table 1 between value rankings of Chinese samples and value rankings for two American national samples, reported by Rokeach in his review of stability of value systems in American culture (Rokeach & Rokeach, 1989). By the logic of validation, correlations between samples of allegedly similar populations should be higher, demonstrated by the high correlations between two American samples in Table 1, than correlations between samples of allegedly different populations. As may be seen in Table 1, the average correlation between Chinese samples is about the same as the average correlation between Chinese samples and American samples.

Katz et al. (1993) measured convergence of results in another way. They compared the consistency of top-ranked and bottom-ranked values of Chinese students in their study and two other studies mentioned above (Feather, 1986; Lau, 1988). They found that the top four values for each Chinese group and the bottom four were idiosyncratic across groups with few exceptions (e.g., salvation was ranked at the bottom by three Chinese groups), and not a single value among the top four and the bottom four were shared by all four groups of Chinese students (Katz et al., 1993, pp. 768–769). In other words, ranking methods do not produce consistent results even for the most important and the least important values for Chinese people.

Table 1

The Spearman's Rank Correlation Coefficients of Rankings of Rokeach's Values Between Samples

Authors	Terminal values (end-states of existence)[a]				Instrumental values (codes of conduct)[b]		
	1	2	3	4	5	6	7
1. Lau (1988)	—	.62	.59	.21	.59	.39	.45
2. Lau & Wong (1992)	.56	—	.80	.50	.58	.45	.47
3. Katz et al. (1993) C	.49	.24	—	.44	.78	.37	.43
4. Katz et al. (1993) A	.37	.39	.48	—	.11	.55	.64
5. Feather (1986)	.68	.40	.51	.38	—	.20	.22
6. American-68 (Rokeach & Rokeach, 1989)	.50	.55	.26	.61	.29	—	.97
7. American-71 (Rokeach & Rokeach, 1989)	.48	.43	.31	.73	.23	.95	—

Note. Correlations for terminal values are presented below the diagonal; correlations for instrumental values above the diagonal. Katz-C refers to the Chinese sample from China and Katz-A refers to the Chinese sample from the US. American-68 refers to the American national sample collected in 1968 and American-71 refers to the American national sample collected in 1971.
[a] Average correlation coefficients for terminal values for within American, within Chinese, and between Chinese & American samples are .95, .45, and .44, respectively.
[b] Average correlation coefficients for instrumental values for within American, within Chinese, and between Chinese and American samples are .97, .52, and .42, respectively.

In this article, we examine the convergence of cross-cultural value comparisons as measured by commonly used ranking methods (e.g., Rokeach's Value Survey) and rating methods (e.g., Schwartz's Value Survey). We also compare four different methods of assessing values—the rating scale, the ranking system, attitude items, and the scenario method—with a validity criterion based on cultural expects' observations. We focus on comparisons between Americans and Chinese, two cultures we are most familiar with and knowledgeable about. Our knowledge of the literature on American and Chinese cultures has caused us to develop some doubts about the accuracy and the validity of value survey methods that simply ask participants "How important is this value to you?"

We also believe that it is possible that one could improve on rating and ranking surveys by using standard attitude questions or by asking for responses to scenarios. The scenario method seems particularly promising because by asking participants questions with a specific context—a hypothetical social situation with various behavioral options—we might be able to avoid the problems of cultural differences in constructing value meanings, as well as the problems of relativity of social comparison-based and deprivation-based value judgments. The reasons are simple: First, a specific situation has already defined what value is being measured so that participants would not have to generate their own conception and mental representation of the value in the abstract. Secondly, if this situation is believed to reflect the same value according to experts' knowledge about both cultures being compared, then it could also avoid the problem of the values having different meanings in two different cultures or languages. Thirdly, the behavioral scenario method that we are suggesting only asks about people's behavioral preference in a hypothetical situation, not their judgments about the importance of an abstract concept, which should largely avoid the problem of social comparison-based and deprivation-based judgments. Finally, because scenarios ask only for preferences, problems related to situational constraints on behavior should be reduced.

Thus, we believe that commonly used ranking and rating methods may have validity problems in describing value differences between American and Chinese people and more context-based attitude and scenario questions might be possible solutions to the problems. On the basis of these concerns, we decided to actually evaluate the convergence of ranking and rating methods and to compare them with the results of attitude and scenario measures. The studies reported here can scarcely be expected to settle any issues concerning validity for value studies, even for the relatively narrow question of validity of value measurement methods for Chinese versus American culture, but we believe they do raise important questions about validation that should be taken into consideration in future research.

Study 1: An Examination of Convergent Validity of the Value Ranking and Rating Methods

It seems clear that Rokeach's Value Survey method may have limitations for describing Chinese value profiles. Is this limitation caused by a particular procedure, namely the ranking method that the Rokeach Value Survey uses? There has been an intense debate in the field of value study about the advantages and disadvantages of the ranking procedure and the rating procedure in cross-cultural comparisons (Feather, 1973; Ng, 1982; Rankin & Grube, 1980; Rokeach, 1973). The major criticism of the ranking method concerns its basic assumptions that everyone has fully developed value systems and a value hierarchy (in which every single value has its unique rank), which might not hold cross-culturally (Ng, 1982; Schwartz, 1992). These scholars advocate the rating method for cross-cultural value comparisons.

In Study 1, we compared results of the ranking method with two different sets of rating data using the rating method as in Schwartz's Value Survey. The ranking data consisted of the composite rankings, that is, the average ranking of Rokeach's values by Chinese participants in previous studies (Feather, 1986; Katz et al., 1993; Lau, 1988; Lau & Wong, 1992). One set of rating data came from a study similar in terms of procedure, sample, and circumstances to the ranking studies by Triandis, McCusker, & Hui (1990). A second set of rating data was collected by ourselves, again using procedures, samples, and circumstances similar to those of the ranking studies. These data sets allowed us to examine the degree of correlation of ranking and rating methods of culture differentiation and the degree of culture differentiation reliability to be obtained with two sets of rating data. Poor reliability of the ranking method among Chinese samples does not preclude the possibility that the ranking method would differentiate Chinese and

American samples in the same way as rating data, thus suggesting that each method might have some validity.

Method

Participants. A total of 100 undergraduate students from Beijing University and 82 undergraduate students from the participant pool at the University of Michigan participated in this study during the fall of 1991. Most of them were first-year students attending introductory psychology courses at the two schools. Both groups were told that they were taking part in a study of perception and that the value data were needed as background information. The gender ratio was matched between these two groups (60% male and 40% female).

Materials. A total of 52 of Schwartz's values was used in this study, of which 29 values were identical with or similar to those in Rokeach's Value Survey (Schwartz & Blisky, 1987; Triandis et al., 1990). The 29 values included 16 of Rokeach's terminal values (excluding happiness and salvation) and 13 instrumental values (excluding cheerful, clean, logical, loving, and obedient). A Chinese version of these 52 values was translated by Kaiping Peng and then translated back from Chinese to English by Chen Chongken, our collaborator at Beijing University.

Procedure. As in the standard rating procedure, participants were asked to rate how much they endorsed each of the 52 values. Each participant answered the question, "What values are important to me as guiding principles in my life, and what values are less important to me?" Participants rated values on a scale from −1 (*rejection of the value*) to 7 (*the value is of supreme importance*).

Results

For the ranking measure, we decided to use composite ranking differences from various survey results as the criterion instead of the results of one survey because the ranking method of the reliability problems for Chinese value survey results. The average rankings of five Chinese samples obtained by previously published studies (Feather, 1986; Katz et al., 1993; Lau, 1988; Lau & Wong, 1992) and the average rankings of two American samples (Rokeach & Rokeach, 1989) were calculated and used as the composite rankings (listed in Table 2, columns 1 and 3, respectively). The Triandis et al. (1990) cross-cultural value survey ratings are listed in Table 2, columns 4

and 5, and our own rating results are listed in columns 6 and 7.

Chen, Lee, and Stevenson (1995) suggested that East Asian students have a moderacy bias that precludes them from moving their responses very far from the middle of the scale. We followed Triandis' recommendation to standardize all the scores (participants' ratings) across values and participants within each culture to eliminate this possible cultural difference in response style on the rating scale (Triandis et al., 1990). Then t tests were used to examine the significance of the mean differences between American and Chinese participants on the standardized scores instead of the raw scores. Results are entirely similar if analyses are conducted on raw scores, however.

Convergence between the ranking procedure and the rating procedure. We checked the percentage of agreement between the ranking differences between cultures and the rating differences between cultures for Rokeach's values. For example, if both measurement results suggested that Americans (A) endorse a given value more than Chinese (C; A > C), we counted it as an agreement. However, if one suggested A > C, the other A < C or A = C, we counted it as a disagreement. For Rokeach's terminal values, the average ratings in the Triandis et al. (1990) study were in the same direction as the composite rankings for only 7 of 16 values; thus the convergence index was .44, indicating trivially less than chance-level agreement. The average ratings in our study were in the same direction as the composite rankings for only 8 out of 16 values; exactly a chance level of agreement. For Rokeach's instrumental values, the average ratings in the Triandis et al. study were in the same directions as the composite rankings for only 5 of 13 values, for a convergence index of .38, somewhat worse than would be expected by chance. The average ratings in our study were in the same direction as the composite ranking for only 3 out of 13 values, marginally significantly worse than chance ($z = 1.94, p < .10$).

The convergence between one rating study and another rating study. We also checked the convergence between the results of our rating study and the results of the Triandis et al. (1990) rating study and found that, for the 52 value comparisons, 27 were rated in the same direction in both studies. The overall convergence index was .52, which is not different from chance. The convergence index for 29 Rokeach's values was .45; for the 23 "other" values, the convergence index was .61.

Table 2

Comparison of Results of Studies Comparing U.S. and Chinese Values

Values	Composite rankings[a]		Triandis, McCusker, & Hui's (1990) ratings[b]		Peng, Nisbett, & Wong's ratings[c]	
	U.S. sample	Chinese sample	U.S. sample	Chinese sample	U.S. sample	Chinese sample
Rokeach's terminal						
An exciting life	7 > 18		4.60 > 3.26***		.04 >	-.60***
A sense of accomplishment (successful)	8 > 10		5.53 < 5.56		.65 >	.39*
World of beauty	11 > 15		3.52 < 3.59		-.31 <	.31***
Freedom	2 > 3		5.42 > 4.07***		.64 <	.73
Mature love	13 > 14		5.41 > 5.11		.68 >	.50
National security	10 > 12		3.76 < 5.07***		-.28 <	.01
Pleasure	16 > 17		4.33 > 4.04		.12 <	.25
Self-respect	4 > 5		5.67 > 5.63		.64 <	.84
Social recognition	12 > 16		4.25 < 5.48**		-.13 <	.15
True friendship	1 > 11		5.73 > 4.93***		.79 >	.58
Wisdom	3 > 6		4.71 < 5.05		29 <	.72***
A comfortable life (wealth)	15 < 9		3.91 > 3.35		-.61 <	-.28
A world at peace	7 < 1		4.21 < 4.23		-.06 >	-.38
Equality	9 < 7		4.43 > 4.33		.22 <	.64*
Family security	5 < 2		5.60 > 4.96		.56 <	.61
Inner harmony	14 < 13		5.25 > 4.48		.68 <	.78
Rokeach's instrumental						
Capable	3 > 9		5.13 < 5.52		.30 <	.40
Imaginative (creative)	15 > 18		3.91 < 4.89**		-.10 <	.42***
Independent	9 > 13		4.94 < 5.00		.22 <	.56*
Intellectual	3 > 15		5.50 > 4.78		.39 >	.30
Responsible	1 > 3		5.26 > 4.81		.54 >	.26
Broadminded	5 = 5		4.44 > 4.26		.00 <	.57**
Polite	14 = 14		4.25 < 4.59		-.05 >	.09
Ambitious	3 < 2		5.06 < 5.07		.26 >	.22
Courage (daring)	7 < 6		3.25 < 3.26		-.60 <	.27
Forgiving	13 < 4		4.27 > 3.70		.02 >	.00
Helpful	11 < 7		4.32 > 3.59		.23 >	-.30***
Honest	6 < 1		5.25 > 4.96		.52 >	-.02***
Self-controlled (self-disciplined)	13 < 10		4.06 < 5.48***		-.02 >	-.74***
Other						
A sense of belonging			5.18 > 1.33***		.48 >	.33
A varied life			4.51 > 3.56**		-.32 <	.14**
Authority			3.07 > 2.30		-.33 >	-.69
Detachment			1.22 < 1.81		-1.19 <	-.91
Devout			3.14 > 2.33***		-.17 >	-1.30***
Enjoying life			5.42 > 3.07***		.31 >	.17
Loyalty			5.31 > 4.41*		.40 >	.14*
Meaning in life			5.08 > 5.00		.40 <	.46
Respect for tradition			2.94 > 2.15		-.64 >	-1.37***
Accepting my position in life			1.96 < 3.30***		-.61 >	-.82
A spiritual life			3.03 < 4.33***		.10 <	.77**
Choosing own goals			5.20 < 5.63		.35 <	.37
Health			5.34 < 5.59		.33 <	.72**
Honoring parents and elders			4.49 < 4.74		-.22 >	-1.10***
Humble			3.38 < 3.67		-.27 >	-.35
Moderate			2.05 < 2.07		-.85 > -	.87

Table 2
Continued

Values	Composite rankings[a]		Triandis, McCusker, & Hui's (1990) ratings[b]		Peng, Nisbett, & Wong's ratings[c]	
	U.S. sample	Chinese sample	U.S. sample	Chinese sample	U.S. sample	Chinese sample
Preserving public image			3.35 < 3.41		−.94 < −.43**	
Protecting the environment			2.82 < 3.26		−.15 > −.33	
Reciprocation of favors			3.73 < 3.85		−.36 < .28	
Social order			3.18 < 4.44***		−.37 < −.21	
Social justice			4.14 < 4.44		−.29 < .08*	
Social power			2.14 < 2.67		−.77 > −1.10	
Unity with nature			2.28 < 3.03*		−.52 < −.01*	

Note. ">," "<," and "=" indicate the value difference between two cultures. In the ranking system, the lower the number the higher the rankings. In the rating system, the lower the number, the lower the ratings.
[a] The convergence index for Triandis et al.'s results and the composite rankings' on Rokeach's terminal values was .44 and the sign test z = −.5, *ns*. The convergence index for Peng et al.'s results and the composite rankings' on the same values was .50, and z = 0, *ns*.
[b] The convergence index for Triandis et al.'s results and the composite rankings' on Rokeach's instrumental values was .38, and z = −.83, *ns*. The convergence index for Peng et al's results and the composite rankings' on the same values was .23, and z = −1.94, *ns*.
[c] The convergence index for Traindis et al.'s results and Peng et al.'s results on Rokeach's values was .45, on other values was .61, and on all the values was .51. The sign test z was −.63, 1.04, .28, respectively, and all of them were not significant.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Study 2: Criterion Validity of Different Value Survey Methods

The low convergence between one set of testing results and other sets of results could be attributed to one of the measures or to both. In Study 2, we examined the possibility that the indications of low validity are not due to the particular criteria we used or the particular samples we studied. We decided to use cultural experts' judgments about the values of Chinese and American peoples as a validity criterion and to use Singapore Chinese, who speak English as a second language, as participants of the study. Furthermore, we added attitude scales and scenario methods in this cross-cultural value comparison to examine possible solutions for the problem of poor convergence of methods we found in Study 1. We expected that attitude scale and scenario method of value measurement might have higher criterion validity than the ranking and rating methods.

Method

Participants. The participants were business school students in Singapore and the United States. Seventy-four ethnic Chinese students were from the National University of Singapore and 64 American students were from the University of Michigan.

Unlike in Study 1, the Chinese participants in this study read an English version of the questionnaire just

like the U.S. participants read. Because one of the official languages in Singapore is English (Lee, 1978), a linguistic explanation for the low criterion validity found in Study 1 should be reduced or eliminated.

Materials. We asked a panel of cultural experts' questions about "Which cultures should endorse the following value concepts more?" These cultural experts were 17 graduate students in Chinese studies at the University of Michigan. Three of them were from China (2 mainland Chinese, 1 Taiwanese) and the rest were from the United States. All the American graduate students had been either in mainland China or Taiwan before and had stayed there for more than 1 year. None of them knew the purpose of the study. No significant differences in judgments about values were found between the experts from China and the experts from the United States ($p > .25$). Their judgments about the direction of the value differences between Americans and Chinese are listed in the first column of Table 3. The numbers inside the parentheses are the percentage of experts agreed on that direction.

We selected thirteen of the values that had the highest agreement among the cultural experts (more than 70% of them agreed on a direction) with a mean of 86% agreement. These experts believed that Americans value an exciting life, choosing own goals, daring, and protecting the environment more than Chi-

Table 3

Standardized Means of Values Measurements Between American (A) and Singapore, Chinese (C) Students by Methods

Value	Expert's judgments	Abstract survey[a]				Attitude scale[b]				Behavioral scenarios[c]			
		Rating		Ranking		Positive		Negative		Scenario 1		Scenario 2	
		U.S.	Singapore	U.S.	Singapore	U.S.	Singapore	U.S.	Singapore	U.S.	Singapore	U.S.	Singapore
An exciting life	A > C (.88)	-1.06 < -.62		.67 > -.15		1.63 > 1.40		-.33 > -.44		.02 > -.38		.20 > .01	
Choosing own goals	A > C (1.00)	-.14 < .91		.84 < 1.19		.61 < .66		-1.04 < -1.01		-.05 > -.97		1.08 > 1.07	
Daring	A > C (.88)	.23 > -.72		-.13 > -1.12		.07 < .70		1.14 > -1.05		.48 > .21		.81 > .36	
Protecting the environment	A > C (.88)	-1.48 < -.36		-2.51 < -.86		-.17 < .83		-.07 < .60		-.32 > -.33		.13 < .89	
Accepting my position in life	A < C (.88)	.18 > -.40		.39 > .37		.30 > .26		-.70 < -.01		.14 < .28		-.74 < -.56	
Authority	A < C (.88)	1.34 > -1.31		-.03 > -.99		-.05 < 1.17		-.79 < -.49		.03 < .76		.69 > -.49	
Honoring parents and elders	A < C (1.00)	.18 < 1.13		-.16 < 1.26		.54 < 1.38		.23 < 2.28		.71 < .81		-.08 < .58	
Humble	A < C (.76)	1.52 > .38		-.38 < .24		1.72 > 1.20		.07 > -.53		.76 < .82		-.84 < -.01	
Inner harmony	A < C (.71)	1.05 < 1.31		1.50 > 1.05		-.37 < .80		.19 > .17		.08 > -.72		-.18 > -.71	
Moderate	A < C (.88)	-1.75 < -1.20		.23 > -.96		-.23 < .80		1.19 > .17		.16 > -.13		-.60 < .37	
Respect for tradition	A < C (.88)	.43 > -1.06		-1.21 > -1.26		-.56 < .04		-1.63 < -.44		.52 < .76		1.27 > .79	
Self-discipline	A < C (.71)	-.17 < 1.55		.63 < 1.47		-.81 < 1.24		.98 > -1.05		-1.12 < -.31		.81 < .99	
Social order	A < C (.88)	-.32 < .40		.12 > -.25		-.39 < 1.36		.81 > .71		.05 < .55		.63 > .47	

Note. All the scores are Z scores; the higher the Z score, the higher the value endorsement (the ranking Z, and negative attitude scores are reversed).
[a] The convergence index for the expert judgments' and the results obtained by using abstract value survey (both rating and ranking methods) was .42, and the sign test $z = -.78$, ns.
[b] The convergence index for the expert judgments and the results obtained by using attitude scale (both positive and negative questions) was .54, and the sign test $z = .39$, ns.
[c] The convergence index for the expert judgments and the results obtained by using behavioral scenario methods was .73, and the sign text $z = 2.35$, $p < .05$.

nese, whereas the Chinese value accepting my position in life, authority, honoring parents and elders, humble, inner harmony, moderate, respect for tradition, self-discipline, and social order more than Americans.

It is almost surely no accident that these value differences were so highly agreed on by our cultural experts. The ethnographic, psychological, historical, and philosophical literature are replete with testimony that Americans hold the particular values indicated by our experts and that Chinese favor the particular values indicated by our experts. Work by Bond (1986), Hsu (1981), Moore (1967), Munro (1985), Nakamura (1964), Northrop (1966), and Triandis (1995) enables us to have confidence in the degree to which our experts' judgments are shared. More specifically, for every individual value, we could easily find scholarly literature to support our experts' claims. Values such as accepting my position in life (see Bloom, 1981), authority (see Singh, Huang, & Thompson, 1962; Ho & Lee, 1974; Meade, 1970; Meade & Whittaker, 1967), honoring parents and elders (see Ho, 1994; Ho & Kang, 1984; Lockett, 1988), "humble" (see Chiu & Kosinski, 1994; Yang, 1970), inner harmony (see Abbott, 1970; Young, 1982), moderate (see Chen et al., 1995), respect for tradition (Abbott, 1976) are traditional Chinese values. The values choosing own goals, daring, and an exciting life are central to the concept of individualism (Markus & Kitayama, 1991; Triandis, Bontempo, Villareal, & Asai, 1988; Triandis et al., 1990). Cultural differences on other values, such as protecting the environment (Goldstein, 1992), self-discipline (see Ho, 1981; Ralston et al., 1992), social order (see Bond & Wang, 1983; Earle, 1969) were also backed by the relevant research. No literature that we are aware of suggests value differences that are inconsistent with our experts' judgments.

In this study, participants were asked to rate and rank the values according to their importance. The rating procedure basically followed Schwartz's Value Survey by using the same 7-point-scale used in Study 1. The ranking procedure followed Rokeach's Value Survey by using a different scale, in which 1 was assigned to the value which participants thought was the most important to them; 2 was assigned to the value which they thought was second most important and so on, until they had ranked all 13 values. No ties were allowed.

Twenty-six attitude questions were also created, 2 for each value. One item measured the value in a positive way, whereas the other measured it in a nega-

tive way. This was done to balance individual differences in response set; for example, the "yes" tendency on Likert-type attitude scales. For example, for measuring participants' value of "accepting my position in life," two attitudinal questions are

1. In general, it is good not to expect too much; that way one does not have to experience frequent disappointments or failures. (Positive item)
2. The higher your goal is, the greater your achievement will be. (Negative item)

We followed two basic principles in the process of creating scenarios: (a) The situation had to be possible in both cultures; (b) the behavioral options had to be specified but would not be seen as absurd or bizarre to the participants from either culture. A total of 26 scenarios were created, 2 for each value. The following is an example of how scenarios measured the value of accepting my position in life:

Scenario 1: Suppose that you have worked very hard in a big company for quite some time, but for reasons that are unclear, you have been passed over for promotion many times. How likely is it that you would prefer to do the following things?
1. Wait for next time and hope for the best.
2. Complain to your supervisors.
3. Threaten your supervisors with a formal protest to the company head.
4. Lower your aims and try to be content in your current position.

Scenario 2: Suppose that you had applied for a job with an organization many years ago, and were offered a less desirable one instead. Nevertheless, you accepted the offer and after many years, you begin to feel satisfied in the job, which you handle very well. You have a nice home, family, and have made many friends during this period of time. Suppose one day, out of the blue, you are offered the job which you originally favored with a big salary increase. The obstacle is that the new job requires you to go overseas for a long period of time. How much would you prefer the following options?
1. Stay on current job.
2. Bargain for salary increase on current job because of this new offer.
3. Go to the new job with some hesitation.
4. Go to the new job without hesitation.

Procedure. The participants were first asked to rate the 13 values in terms of their importance as the "guiding principle for their life," as they did in Study 1. Then they responded to the 26 attitude items (the order had been randomly arranged) on a 7-point scale from *strongly disagree* to *strongly agree.* The instruction for the attitude scale was "Please indicate the

extent to which you agree/disagree with each of the following statements." After they finished this part, participants were asked to rank this list of values according to their importance.

Next, participants were asked to read each scenario and choose the behavior they would prefer if they were in that situation. A 7-point scale from *definitely not* to *definitely* was provided to let participants indicate how likely it was that they would prefer each behavioral option.

Finally, participants were asked to match the scenarios with the values. Participants could correctly match 88% of the values with scenarios, which demonstrated that the behavioral scenarios had considerable face validity to the participants for measuring the values they were intended to measure.

Results

As in Study 1, participants' raw scores were converted to standardized z scores though results are entirely similar when it is raw scores that are analyzed. The z scores of rankings, and of negative attitude items and behavioral options, were reversed so as to be consistent with the other scores in the direction of value endorsement. The scenario z scores were means of participants' responses to the two behavioral options in each scenario. If there were multiple options in a scenario, the ratings for the most preferred behavioral option for Chinese and the most preferred option for Americans were averaged.

Next, a convergence index was calculated for each method based on the number of agreements between the experts' judgments and the value survey results obtained by each method. The significance level for each convergence index was assessed by nonparametric sign test zs as in Study 1. This validity analysis yielded the following results.

Criterion validity of the ranking and rating procedures of value surveys. We found that the standard methods of value surveys, both ranking and rating procedures, had the lowest criterion-related validity. The convergence index was .42, less than a chance level of agreement with experts' opinion. The criterion validities of the ranking and rating procedures were nearly the same, although the rating procedure had a trivially higher convergence index (.45 vs. .39).

Criterion validity of the attitude scale. The attitude scale measurements produced a convergence index of .54, which indicates that only slightly more than half of the results were consistent with the ex-

perts' judgments (14 out of 26). This of course scarcely exceeds the chance level of .50.

Criterion validity of the scenario method. As expected, the behavioral scenario method had the highest convergence index. Eighteen out of 25 pairs were in the same direction as the judgments of the experts' (one scenario had the same scores for both groups, and we did not include this one in the calculation). The sign test $z = 2.20$ ($p < .05$) for the convergence index of .72, shows that this method was successful in predicting the judgments of cultural experts.

Discussion

Limitation of Commonly Used Value Survey Methods

The results suggest that the method of asking participants to rate or rank the importance of a set of values and then aggregating these ratings or rankings for comparison cross-culturally may not provide a good picture of true value differences. There has been an intense debate about the advantages and disadvantages of the ranking procedure and the rating procedure in cross-cultural value comparisons (Feather, 1973; Ng, 1982; Rankin & Grube, 1980; Rokeach, 1973). Our research suggests that the agreement between these two procedures is not very high, indeed the agreement of each procedure with itself may sometimes not be very high, and that neither method guarantees high criterion validity.

We were not totally surprised by the fact that the attitude items for the value measurement did not perform significantly better than chance. Nisbett and Cohen studied violence in different U.S. regional cultures, and found that no cultural differences existed between northerners and southerners regarding how much they reported disliking violence and physical conflict in any of the abstract value surveys or attitude studies administered. However, with a survey having specific contextual-based questions or scenarios (e.g., concerning defense of one's propriety, protecting one's female companions, or responding with violence to personal insults), they found that Southerners were more likely to choose violent behavioral options than Northerners. These results are consistent with the actual homicide rates in the North and the South of the United States (Nisbett, 1993; Cohen & Nisbett, 1994). A lesson of the great debate about the attitude–behavioral inconsistency in social psychology is that when the attitude questions are nonspecific, not experience-related, and too general, they will not predict

people's behaviors (see reviews by Fazio & Zanna, 1981; Tesser & Shaffer, 1990). The sorts of questions in value surveys are of course even more broad and abstract than those used in most attitude–behavioral studies.

A Case for Experimentation: Advantages of the Scenario Method

It appears that the low criterion validity of commonly used value survey methods might be avoided by using the behavioral scenario method. This method may be useful because it can reduce "noise," such as differences between cultures in the interpretation of the meaning of value terms and such problems as the relativity of social comparison-based judgments and deprivation-based preferences. For example, Americans might say they can "accept their positions in life" when you ask them their preference in an abstract way. However, if there is a chance for advancement with some risks, they may be likely to take it or encourage others to do so, in which case they actually do not value "accepting my position in my life" very much.

One reasonable question about the current studies might be the following: Because participants' self-reports about values are such unreliable and invalid indications for cross-cultural value comparison, why not just observe people's behaviors in their real lives? Why bother to ask their behavioral preferences in a hypothetical situation? The answers to these questions are based on one of the fundamental lessons of social psychology, which is the power of small situational factors to influence people's behavior. For people who live in different cultures, setting up experimental situations having identical meaning, or even establishing observational settings that have the same meaning, can be extremely difficult. It is hard enough to ensure even within a culture that such details as perceived degree of choice, attributed status of the experimenter, and so on, are equated across populations and laboratories. To be sure of such matters, which can be crucial to demonstrating a phenomenon, is far more difficult when comparisons are being made across cultures. In a sense, the scenario method may be like a behavioral observation in controlled conditions where unwanted situational factors are minimized. Another difficulty with behavioral observation is the enormous labor involved in creating situations and collecting data for a sample of behavioral measures for each of 40–50 values comparing even two cultures, let alone many cultures.

Finally, one may ask about the usefulness of doing psychological research at all if cultural experts' opinions are to be the ultimate criterion. There are several answers to this question. First, the opinion of experts is never meant, in validation research, to be unappealable. If investigators' intuitions, or relevant data, suggest that the experts' opinion (and the larger nomological net in which these are situated) is in error, such hypotheses are always testable in principle. The history of clinical diagnosis is full of instances where expert views had to yield to evidence from the laboratory or the field. Second, there may be no expert opinion, or there may exist only highly divided expert opinion, in the case of some cultures or subcultures. In such cases, a highly valid criterion is missing, and research comparing cultures can be carried out only by means other than inquiring of experts. Obviously it is preferable to use means that have proved valid in contexts where clear validity criteria do exist. Third, objective and replicable measures of values are needed because cultural values can change rapidly and must be studied by social scientists, a point made convincingly by research by Inglehart (1990).

Conclusion and Qualifications

We do not wish to be in the position of overgeneralizing the results from the present study. It does not show that rating or ranking methods cannot have cross-cultural validity, nor does it show that attitude items cannot be generated that are capable of predicting expert judgments, nor does it show that well-constructed scenarios will always give more valid results than other methods—even for the single pair of cultures we examined. The study should instead be regarded as an "existence proof" of narrow versions of each of these points: Ranking methods may correlate only modestly with themselves and may correlate little or not at all with rating methods, and attitude items may not predict expert judgments whereas scenario methods may. But even these relatively narrow points should serve as a warning that validity for common methods of cross-cultural comparisons of values cannot be assumed but must be proved.

References

Abbott, K. A. (1970). Harmony and individualism. Taipei, Taiwan: Oriental Cultural Service.

Abbott, K. A. (1976). Culture change and the persistence of the Chinese. In G. A. De Vos (Ed.), Response to change:

Society, culture and personality (pp. 74–103). New York: Van Nostrand, Reinhold.

Allport, G., Vernon, P., & Lindzey, G. (1951). *Study of value* (2nd ed.). Boston: Houghton Mifflin.

American Psychological Association. (1974). *The standard for educational and psychological testing.* Washington, DC: Author.

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.

Appleton, S. (1970, April–June). Surveying the values of Chinese college students. *Asian Forum, 75–88.*

Appleton, S. (1979). Sex, values, and change in Taiwan. In R. W. Wilson, A. A. Wilson, & S. L. Greenblatt (Eds.), *Value change in Chinese society* (pp. 185–202). New York: Praeger.

Bloom, A. H. (1981). *The linguistic shaping of thought: A study in the impact of language on thinking in China and the West.* Hillsdale, NJ: Erlbaum.

Bond, M. H. (1986). *The psychology of the Chinese people.* Hong Kong: Oxford University Press.

Bond, M. (1988). Finding universal dimensions of individual variation in multicultural studies of values: The Rokeach and Chinese value surveys. *Journal of Personality and Social Psychology, 55,* 1009–1015.

Bond, M. H., & Wang, S. H. (1983). Aggressive behavior in Chinese society: The problem of maintaining order and harmony. In A. P. Goldstein & M. Segall (Eds.), *Global perspectives on aggression* (pp. 58–74). New York: Pergamon Press.

Braithwaite, V., & Law, H. (1985). Structure of human values: Testing the adequacy of the Rokeach Value Survey. *Journal of Personality and Social Psychology, 49,* 250–263.

Bratz, K. W., & Levine, E. S. (1978). Childrearing by black parents: A description and comparison of Anglo and Chicano parents. *Journal of Marriage and Family Living, 40,* 707–719.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 61–105.

Chen, C., Lee, S. Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science, 6,* 170–175.

Chiu, R. K., & Kosinski, F. A. (1994). Is Chinese conflict-handling behavior influenced by Chinese values? *Social Behavior and Personality, 22,* 81–90.

Cohen, D., & Nisbett, R. (1994). Self-protection, insults and the culture of honor: Explaining southern homicide. *Journal of Personal and Social Psychology Bulletin, 20,* 551–567.

Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.

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

Crosby, F. (1976). A model of egotistical relative deprivation. *Psychological Review, 83,* 85–113.

Earle, M. J. (1969). A cross-cultural and cross-language comparison of dogmatism scores. *Journal of Social Psychology, 79,* 19–24.

Fazio, R., & Zanna, M. (1981). Direct experience and attitude–behavior consistency. In L. Berkowitz (Ed.), *Advances in experimental psychology* (Vol. 14, pp. 161–202). New York: Academic Press.

Feather, N. (1973). The measurement of value: Effects of different assessment procedures. *Australian Journal of Psychology, 25,* 221–231.

Feather, N. (1980). Similarity of values systems within the same nation: Evidence from Australia and Papua New Guinea. *Australian Journal of Psychology, 32,* 17–30.

Feather, N. (1986). Value systems across cultures: Australia and China. *International Journal of Psychology, 38,* 297–308.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117–140.

Fuligni, A., & Stevenson, H. (1994). *Time-use and academic achievement among Chinese, Japanese, and American high school students.* Unpublished manuscript, University of Michigan—Ann Arbor.

Goldstein, S. M. (1992). *China at the crossroads: Reform after Tiananmen.* New York: Foreign Policy Association.

Hare, B. B., & Castanell, L. A., Jr. (1985). No place to run, no place to hide: Comparative status and future prospects for black boys. In M. B. Spencer, G. K. Brookins, & W. R. Allen (Eds.), *Beginnings: The social and affective development of black children* (pp. 117–130). Hillsdale, NJ: Erlbaum.

Ho, D. Y. F. (1981). Traditional patterns of socialization in Chinese society. *Acta Psychologica Taiwanica, 23*(2), 81–95.

Ho, D. Y. F. (1994). Filial piety, authoritarian moralism, and cognitive conservatism in Chinese societies. *Genetic, Social, & General Psychology Monographs, 120,* 347–365.

Ho, D. Y., & Kang, T. K. (1984). Intergenerational comparisons of child-rearing attitudes and practices in Hong Kong. *Developmental Psychology, 20,* 1004–1016.

Ho, D. Y. F., & Lee, L. Y. (1974). Authoritarianism and attitudes toward filial piety in Chinese teachers. *Journal of Social Psychology, 92,* 305–306.

Hofstede, G. (1980). *Culture's consequences: International*

differences in work-related values. Beverly Hills, CA: Sage.

Homer, P. M., & Kahle, L. R. (1988). A structural equation test of the value-attitude-behavior hierarchy. *Journal of Personality and Social Psychology, 54,* 638–646.

Hsu, F. L. K. (1981). *American and Chinese: Passage to differences* (3rd ed.). Honolulu: University of Hawaii Press.

Inglehart, R. (1990). *Culture shift in advanced industrial society.* Princeton, NJ: Princeton University Press.

Iwata, O. (1991). A cross-cultural comparison of socially appropriate behavior: A comparison between Japanese and United States undergraduates. *Psychologia, 34*(3), 149–156.

Kahle, L. (1984). *Attitudes and social adaptation.* Elmsford, NY: Pergamon Press.

Kahneman, D., & Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Katz, B., Juni, S., & Shope, C. (1993). The values of Chinese students: At home and abroad. *International Journal of Psychology, 28,* 761–773.

Keats, D. (1986). Using the cross-cultural method to study the development of values. *Australian Journal of Psychology, 38,* 297–308.

Kluckhohn, F. R. (1951). Values and value orientations in the theory of action. In T. Parsons & E. A. Shils (Eds.), *Toward a general theory of action.* Cambridge: Harvard University Press.

Lau, S. (1988). Value orientations of Chinese students in Hong Kong. *International Journal of Psychology, 23,* 583–596.

Lau S., & Wong, S. (1992). Value and sex role of Chinese adolescents. *International Journal of Psychology, 27*(3), 3–17.

Lee, K. (1978). *Bilingualism in our society.* Singapore City, Singapore: Ministry of Culture of Singapore, Publicity Division.

Lockett, M. (1988). Culture and the problems of Chinese management. *Organization Studies, 9,* 475–496.

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

Meade, R. D. (1970). Leadership studies of Chinese and Chinese-Americans. *Journal of Cross-Cultural Psychology, 1,* 325–332.

Meade, R. D., & Whittaker, J. O. (1967). A cross-cultural study of authoritarianism. *Journal of Social Psychology, 72,* 3–7.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educa-* *tional measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Moore, C. A. (1967). *The Chinese mind: Essentials of Chinese philosophy and culture.* Honolulu, HI: East-West Center Press.

Morris, C. W. (1956). *Varieties of human value.* Chicago: University of Chicago Press.

Munro, D. J. (1985). *Individualism and holism: Studies in Confucian and Taoist values.* Ann Arbor: University of Michigan, Center for Chinese Studies.

Nakamura, H. (1964). *Ways of thinking of Eastern people: India, China, Tibet, Japan.* Honolulu, HI: East-West Center Press.

Nisbett, R. (1993). Violence and U.S. regional culture. *American Psychologist, 48,* 441–449.

Ng, S. H. (1982). Choosing between the ranking and rating procedures for the comparison of value across cultures. *European Journal of Social Psychology, 12,* 169–172.

Ng, S. H., Akhtar-Hossain, A. B. M., Ball, P., Bond, M. H., Hayashi, K., Lim, S. P., O'Driscoll, M. P., Sinha, D., & Yang, K. S. (1982). Human values in nine countries. In P. Rath, H. S. Asthana, D. Sinha, & J. B. P. Sinha (Eds.), *Diversity and unity in cross-cultural psychology* (pp. 196–205). Lisse, The Netherlands: Swets & Zeitlinger.

Northrop, F. S. C. (1966). *The meeting of East and West: An inquiry concerning world understanding.* New York: Collier-Macmillan.

Peng, K. (1989). *Psychological testing: The theories and practices.* Beijing, China: Huaxia Press.

Peng, K., & Chen, Z. (1989). Chinese value orientation in university students. *Acta Psychologica Sinica, 21,* 149–155.

Pitts, R. E., & Woodside, A. G. (Eds.). (1984). *Personal values and consumer psychology.* Lexington, MA: Heath.

Ralston, D. A., Gustafson, D. J., Elsass, P. M., & Cheung, F. (1992). Eastern values: A comparison of managers in the United States, Hong Kong, and the People's Republic of China. *Journal of Applied Psychology, 77,* 664–671.

Rankin, W., & Grube, J. (1980). A comparison of ranking and rating procedures for value system measurement. *European Journal of Social Psychology, 10,* 233–246.

Rokeach, M. (1973). *The nature of human value.* New York: Free Press.

Rokeach, M. (1979). *Understanding human values.* New York: Free Press.

Rokeach, M., & Rokeach, S. (1989). Stability and change in American value priorities, 1968–1981. *American Psychologist, 44,* 775–784.

Ross, L., & Nisbett, R. (1991). *The person and the situation: Perspectives of social psychology.* Philadelphia: Temple University Press.

Schwartz, S. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology, 25,* 1–65.

Schwartz, S., & Blisky, W. (1987). Toward a psychological structure of human values. *Journal of Personality and Social Psychology, 53,* 550–562.

Schwartz, S., & Blisky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology, 58,* 878–891.

Schwartz, S., & Inbar-Saban, N. (1988). Value self-confrontation as a method to aid in weight loss. *Journal of Personality and Social Psychology, 54,* 396–404.

Singh, P. N., Huang, S., & Thompson, G. C. (1962). A comparative study of selected attitudes, values, and personality characteristics of American, Chinese and Indian students. *Journal of Social Psychology, 57,* 123–32.

Stevenson, H., Lee, S., & Graham, T. (1993). Mathematics achievement of Chinese, Japanese & American children. *Science, 259,* 53–38.

Stevenson, H., Lee, S., Chen, C., Stigler, J. Hsu, C., & Kitamura, S. (1990). Contexts of achievement: A study of American, Chinese, and Japanese children. *Monographs of the Society for Research in Child Development, 55* (1–2, Serial No. 221).

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers.* San Francisco: Jossey-Bass.

Tesser, A., & Shaffer, D. (1990). Attitudes and attitude change. *Annual Review of Psychology, 41,* 479–523.

Tetlock, P. E. (1986). A value pluralism model of ideological reasoning. *Journal of Personality and Social Psychology, 50,* 819–827.

Toler, C. (1975). The personal values of alcoholics and addicts. *Journal of Clinical Psychology, 31,* 554–557.

Triandis, H. C. (1995). *Individualism & collectivism.* Boulder, CO: Westview Press.

Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M. (1988). Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships. *Journal of Personality and Social Psychology, 54,* 323–338.

Triandis, H., McCusker, C., & Hui, C. H. (1990). Multimethod probes of individualism and collectivism. *Journal of Personality and Social Psychology, 59,* 1006–1020.

Williams, R. M., Jr. (1968). Values. In E. Sills (Ed.), *International encyclopedia of the social sciences* (pp. 283–287). New York: Macmillan.

Williams, R. M., Jr. (1975). Relative deprivation. In L. Coser (Ed.), *The idea of social structure: Papers in honor of Robert K. Merton.* New York: Harcourt Brace Jovanovich.

Yang, K. S. (1970). Authoritarianism and evaluation of appropriateness of role behavior. *Journal of Social Psychology, 80,* 171–181.

Yang, K. (1986) Chinese personality and its change. In M. Bond (Ed.), *The psychology of the Chinese people* (pp. 115–131). Hong Kong: Oxford University Press.

Young, L. W. L. (1982). Inscrutability revisited. In J. J. Gumperz (Ed.), *Language and social identity* (pp. 72–84). Cambridge, England: Cambridge University Press.

Dannemiller Appointed Editor of *Developmental Psychology*, 1999–2004

The Publications and Communications Board of the American Psychological Association announces the appointment of James L. Dannemiller, PhD, University of Wisconsin, as editor of *Developmental Psychology* for a 6-year term beginning in 1999.

Effective January 1, 1998, manuscripts should be directed to

James L. Dannemiller, PhD
Developmental Psychology Journal Office
Room 555 Waisman Center
University of Wisconsin—Madison
1500 Highland Avenue
Madison, WI 53706-1611
email: jldannem@facstaff.wisc.edu