A Review of Cross-Cultural Methodologies for Organizational Research: A Best-Practices Approach

BRYAN S. SCHAFFER
CHRISTINE M. RIORDAN
University of Georgia

Cross-cultural studies that use self-report instruments can present researchers with a variety of challenges. This article reviews the organizational research literature between the years of 1995 and 2001 to identify common practices prevalent in this type of research. Key methodological issues are examined within the context of a three-stage framework: (a) the development of the research question, (b) the alignment of the research contexts, and (c) the validation of the research instruments. This examination serves as a basis for the identification of best-practice recommendations for cross-cultural researchers.

Keywords: cross-cultural; organizational studies; methods; best practices

International perspectives are prevalent in today’s study of organizations. As business practices become more global, many theoretical constructs commonly used in domestic research are being applied in new cross-cultural arenas. This trend has prompted researchers to highlight potential methodological issues associated with conducting this type of research (e.g., Cheung & Rensvold, 1999; Riordan & Vandenberg, 1994). Some of these issues include whether the researchers take an emic or etic perspective, whether they treat or define culture appropriately in the development of their research questions, and whether they establish equivalence in their selection of samples, their administration of surveys, and in their operationalization of constructs across different cultural groups. If researchers ignore the methodological issues common to cross-cultural research, they risk interpreting findings that may actually be meaningless, inconclusive, or misleading.

The purposes of this article are threefold. First, we provide a review of the important methodological issues involved in the use of self-report instruments in cross-cultural research. Other researchers have suggested how such issues can be threats to validity in a variety of field-research settings (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979; Cook, Campbell, & Peracchio, 1990). Our goal here is to relate these threats specifically to cross-cultural settings.
Second, we identify common methodological practices within a large sample of cross-cultural studies. We discuss these practices and provide examples for clarification. Finally, based on the identification of common practices, we propose key best practices for conducting cross-cultural research with self-report instruments. The analysis of previous cross-cultural studies and the examples provided throughout the article are detailed and clear. In this sense, our review should be useful to current cross-cultural researchers as they pursue their own studies or review research for possible publication.

Our review extends previous reviews on cross-cultural research (e.g., Adler, 1983; Boyacigiller & Adler, 1991; Cavusgil & Das, 1997; Cheng, 1989, 1994; Poortinga, 1989; Triandis, 1994a) in many ways. As outlined in the purposes above, we specifically base our recommendations on key methodological issues associated with survey research. This focus differs from some other reviews that have highlighted trends or that have addressed broader, more overarching issues related to global management research in general. In addition, the sample we use provides a basis for offering a fresh perspective on the current state of cross-cultural management research. Our discussion, which follows a three-stage research framework, highlights important methodological issues that can arise in all phases of cross-cultural research, including initial theorizing, data collection, and the interpretation of results.

Literature Review

To identify common practices, we reviewed eight academic management journals, and the Journal of Cross-Cultural Psychology, to obtain a sample of 210 cross-cultural research studies published between 1995 and 2001. Our criteria for selection were based on previous ratings of management and organizational research journals (see Gomez-Mejia & Balkin, 1992; Johnson & Podsakoff, 1994; Zickar & Highhouse, 2001) along with our own assessment of where these types of studies are predominantly published. Although our sample is not exhaustive, it is fairly representative of cross-cultural organizational research and therefore provides an appropriate basis for identifying common practices and making best practice recommendations (see Table 1 for a list of the journals, and see the appendix for a list of the articles included in the sample).
We considered a study cross-cultural if it focused on the comparison of survey responses gathered across different cultural samples. In addition, a single-culture research study was viewed as cross-cultural if it in some way accounted for differences between cultural settings. This is in line with J. G. Miller (1997), who maintained that analyses of single cultures can still be in the tradition of cross-cultural research. The goal of such research is to gain an understanding of differences between cultures, even if it does not directly compare data across the cultures (Poortinga, 1997). For example, studies assessing the appropriateness of using U.S.-based survey instruments in non-U.S. settings would fall into this category (e.g., Iverson & Maguire, 2000).

A three-stage framework was developed for comparing and evaluating the various methodologies across previous cross-cultural studies. Stage 1 involves the development of the theoretical question and the operationalization of culture within the context of the study. Researchers at this stage should be concerned with the assumptions they make concerning the applicability of constructs across cultures. In addition, they should be aware of instances where other cultural indicators in the research setting, besides country or nation, might allow them to make more precise and relevant hypotheses about cultural differences. Stage 2 involves the alignment of the research contexts, where the primary concern is in making sure that procedures and methods, including sampling techniques, are applied consistently across different cultural groups. A goal for researchers in this stage would be to minimize the effects of differences or inconsistencies that are not relevant to the main purposes of the study. Finally, Stage 3 deals with instrument validation, where researchers are often faced with the issue of whether to develop a new survey instrument or adapt an existing one. The main concern here is in establishing or maintaining the construct validity of the scales used within the cross-cultural study.

Table 2 presents the common practices identified in each stage, along with frequencies representing each practice’s prevalence among the overall sample of studies. Our comparisons and evaluations of the different studies, in the context of these three stages, served as a basis for the identification of some sound cross-cultural research practices. The review that follows is presented in the order of these stages, and within each stage, a discussion of important methodological issues precedes the presentation of these best practices.

**Stage 1: Development of Cross-Cultural Research Questions**

The development and focus of the cross-cultural research question has important implications for further design and measurement within a study. Two issues are particularly relevant for this stage. Researchers must establish whether their studies will have an emic or etic perspective, and they must also determine the way in which they will define or treat culture.

**The Emic-Etic Issue**

*Emic approach.* The emic approach focuses on examining a construct from within a specific culture and understanding that construct as the people from within that culture understand it (Gudykunst, 1997). Much of the research based in the United States, for example, is emic in the sense that it examines work-related issues as they apply to
American employees (such emic research is common in other countries as well). Emic studies can be considered cross-cultural when they take into account, either implicitly or explicitly, other cultures. This type of research often takes constructs, theories, or measures that have been developed in one culture (typically the United States) and

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Yes</th>
<th>No</th>
<th>Could Not Determine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Practices Identified From Sample (N = 210)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development of the cross-cultural research question</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emic/etic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used emic approach?</td>
<td>12</td>
<td>6</td>
<td>196 94 1 209</td>
</tr>
<tr>
<td>Used etic approach?</td>
<td>196</td>
<td>94</td>
<td>12 6 1 209</td>
</tr>
<tr>
<td>Used imposed etics/pseudo etics?</td>
<td>165</td>
<td>79</td>
<td>43 21 1 209</td>
</tr>
<tr>
<td>Used emic-etic (derived etic) approach?</td>
<td>31</td>
<td>15</td>
<td>177 85 1 209</td>
</tr>
<tr>
<td>Operationalizing culture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used country as a proxy for culture?</td>
<td>157</td>
<td>79</td>
<td>40 20 1 198</td>
</tr>
<tr>
<td>Other delimiters besides country?</td>
<td>21</td>
<td>10</td>
<td>178 89 2 201</td>
</tr>
<tr>
<td>Hofstede’s cultural value dimensions part of the study?</td>
<td>83</td>
<td>41</td>
<td>119 59 0 202</td>
</tr>
<tr>
<td>If yes, was country used as a proxy for these dimensions?</td>
<td>44</td>
<td>53</td>
<td>39 47 0 83</td>
</tr>
<tr>
<td>Alignment of the research contexts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample differences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ignored demographic differences</td>
<td>25</td>
<td>15</td>
<td>109 64 36 170</td>
</tr>
<tr>
<td>Ignored differences in environmental/industry characteristics</td>
<td>18</td>
<td>11</td>
<td>128 75 24 170</td>
</tr>
<tr>
<td>Ignored differences in experience levels</td>
<td>25</td>
<td>15</td>
<td>109 64 35 169</td>
</tr>
<tr>
<td>Reported details about samples</td>
<td>122</td>
<td>72</td>
<td>46 27 2 170</td>
</tr>
<tr>
<td>Matched samples</td>
<td>94</td>
<td>56</td>
<td>45 27 29 168</td>
</tr>
<tr>
<td>Statistically controlled for differences</td>
<td>62</td>
<td>37</td>
<td>77 46 29 168</td>
</tr>
<tr>
<td>Survey administration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Described procedures for ensuring equivalent survey formats?</td>
<td>57</td>
<td>33</td>
<td>115 67 0 172</td>
</tr>
<tr>
<td>Described procedures for ensuring equivalent survey timing?</td>
<td>42</td>
<td>24</td>
<td>130 76 0 172</td>
</tr>
<tr>
<td>Described procedures for ensuring equivalent levels of rapport?</td>
<td>13</td>
<td>8</td>
<td>158 92 0 171</td>
</tr>
<tr>
<td>Provided details about procedural equivalence in survey administration?</td>
<td>63</td>
<td>36</td>
<td>110 64 0 173</td>
</tr>
<tr>
<td>Validation of the research instrument</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic equivalence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used back-translation?</td>
<td>108</td>
<td>62</td>
<td>31 18 34 173</td>
</tr>
<tr>
<td>Avoided the use of common U.S. phrases/words in survey instruments?</td>
<td>77</td>
<td>42</td>
<td>8 4 99 184</td>
</tr>
<tr>
<td>Described procedures for ensuring semantic equivalence (mentioned semantic equivalence)?</td>
<td>78</td>
<td>42</td>
<td>107 58 0 185</td>
</tr>
<tr>
<td>Conceptual/scaling equivalence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used covariance structure analysis?</td>
<td>30</td>
<td>17</td>
<td>138 77 11 179</td>
</tr>
<tr>
<td>Used item response theory?</td>
<td>3</td>
<td>2</td>
<td>166 93 10 179</td>
</tr>
<tr>
<td>Other method(s)?</td>
<td>17</td>
<td>9</td>
<td>150 84 12 179</td>
</tr>
<tr>
<td>Described procedures for ensuring equivalence?</td>
<td>45</td>
<td>25</td>
<td>134 75 0 179</td>
</tr>
</tbody>
</table>
adapts them for use within other cultures. In addition, these studies commonly include narrative comparisons between theories or findings in one culture and previous research in other cultures.

An important issue for researchers to consider when using an emic approach is that shared frames of references may not exist across cultures (Ronen & Shenkar, 1988). Since this approach studies behavior from within a single culture, importance is given to understanding insiders’ viewpoints and their cognitive thinking patterns within the particular setting (Weick, 1979). Thus, the unique features of a particular culture are incorporated into the theory, hypotheses, measurement, and analyses, and generalizability across cultures may be limited.

In our sample of cross-cultural studies, only 12 (6%) used an emic analysis. A recent study examining the organizational citizenship behaviors of contingent workers in Singapore exemplifies the use of this emic cross-cultural approach (see Van Dyne & Ang, 1998). These researchers wanted to identify a sample in which contingent employees worked as temporaries on a voluntary basis. Thus, the study was conducted in Singapore, where the unemployment rate is persistently low and where those who want to work as regular employees have little difficulty finding jobs. The authors claimed that examining temporary workers in the United States would be different because in many cases individuals working in the United States as temporary employees do so involuntarily. There was an implicit comparison between two cultures in this research design, despite the fact that the analyses were conducted in one culture. Furthermore, the researchers adapted constructs and theory developed in the United States to take into account the unique aspects of the Singaporean culture within their hypotheses and analyses. For other examples of the emic approach, see J. S. Miller, Hom, and Gomez-Mejia (2001) and Iverson and Maguire (2000).

**Etic approach.** Whereas the emic approach captures important aspects of the particular culture under study, the etic approach employs broader comparative analyses involving two or more cultures. Specifically, etic cross-cultural research involves developing an understanding of a construct by explicitly comparing it across cultures using predetermined characteristics. In our sample, virtually all of the studies (94%) were comparative and were conducted with an etic approach. For example, in examining fairness reactions to personnel selection techniques, Steiner and Gilliland (1996) administered their survey to French and U.S. respondents and proceeded with some comparative analyses. The main assumption with this type of research is that shared frames of references exist across culturally diverse samples. Thus, key constructs (and construct measurement) are usually applied to all samples in the same way, ultimately allowing for more generalizability (Ronen & Shenkar, 1988). In this sense, measurement criteria in an etic approach are often viewed as common absolutes that can be applied across cultures (Berry, 1979; Hesketh & Rounds, 1995).

Cross-cultural researchers often use the etic approach because of certain features that are thought to facilitate the research process. For example, differing events around the world might be viewed with broader perspectives, providing a basis for which similarities and differences can be recognized (Berry, 1990). Researchers may also consider this approach to be the most practical in terms of financial limitations and time pressures. If constructs are perceived to be generalizable across cultures, resource expenditures will likely decrease because researchers will not have to study the emic aspects of each culture individually.
Despite such advantages, researchers often inappropriately use the etic approach to make cross-cultural comparisons without fully taking into account some relevant culture-specific emics. This failure to consider such emic factors, along with the assumption that key constructs exist equally across all cultures, has been labeled “imposed etics,” or “pseudo etics” (Berry, 1990). In other words, an etic shortcut is utilized when perhaps a more thorough emic analysis would have been warranted. The imposed-etic approach is problematic because it can influence researchers to generalize comparative findings to other settings when true differences or similarities may actually be due to underlying cultural factors. This practice has been recognized as being fairly common in cross-cultural research (see Ongel & Smith, 1994), and in our review, we observed it in most of the studies we examined (79%).

For example, Gabrielidis, Stephan, Ybarra, Pearson, and Villareal (1997) compared the conflict resolution styles of college students from the United States and Mexico. Their measures included survey items assessing the independence-interdependence of the self. One item from the survey was, “It is important to me to maintain harmony in the group.” We classified this study as imposed etics because there is an underlying assumption that respondents from both cultures shared the same frame of reference when conceptualizing such terms as harmony. However, no prior analysis was conducted to examine more context-specific meanings of the construct (see also Aryee, Fields, & Luk, 1999, who used imposed etics when they examined the cross-cultural applicability of a work-family interface model).

**Best practices.** As a best-practice approach, we suggest a combined emic-etic or a derived etic approach when making cross-cultural comparisons in organizational research. Rather than identifying emic dimensions from one culture and simply applying those dimensions to the other culture(s) in a study, a derived etic approach requires researchers to first attain emic knowledge (usually through observation and/or participation) about all of the cultures in the study (Berry, 1990; Cheung, Conger, Hau, Lew, & Lau, 1992). This allows researchers to put aside their culture biases and to become familiar with the relevant cultural differences in each setting. When this is done, it may then be possible to make cross-cultural links between the emic aspects of each culture. Although some common dimensions will emerge in all cultures, some dimensions may emerge in only one of the cultures (Cheung et al., 1992). Only where there are observed commonalities can cross-cultural comparisons appropriately be made. In our sample, 15% of the studies used this best-practice approach.

Farh, Earley, and Lin (1997) used the derived etic approach in their examination of the relationship between citizenship behavior and organizational justice in a Chinese context. Three independent Chinese samples were used to develop a Chinese organizational citizenship behavior scale. The organizational citizenship behavior dimensions that emerged from this culture-specific process were analyzed in conjunction with the dimensions that have been identified here in the United States (e.g., Organ, 1988; Podsakoff, MacKenzie, Moorman, & Fetter, 1990). By identifying organizational citizenship behavior dimensions that emerged in both the Western and the Chinese scale, Farh et al. were able to isolate etic aspects of citizenship behavior that could be appropriately applied across cultures. The specific etic dimensions of organizational citizenship behavior were civic virtue, altruism, and conscientiousness. The emic dimensions were sportsmanship, courtesy (both of which emerged only in the Western context),
interpersonal harmony, and protecting company resources (both of which emerged only in the Chinese context) (see Farh et al., 1997, Table 2, p. 429).

In a similar fashion, Gelfand et al. (2001) used a derived etic approach when they examined cognitive representations of conflict in the United States and Japan. Their results revealed some universal dimensions of conflict construal (across both contexts) but also indicated that there were some culture-specific dimensions uniquely applicable to either the United States or Japan.

These procedures are in line with Church and Katigbak’s (1988) suggestion that researchers should search for universal (or derived etic) components of constructs by assessing whether the emerging dimensions are unique to one culture, comparable across cultures, or overlapping. Triandis (1992) effectively described this type of approach in the context of studying individualism and collectivism. His proposed steps are as follows:

1. Begin with a theoretical framework and decide what specific constructs are to be studied.
2. Engage in idea sharing across different cultures about the constructs, with researchers from all cultures working together (emics).
3. Generate items and have samples of convenience respond to all items. Isolate etic dimensions during this step, for example, factors that look alike (items that are determined to have different meanings across different cultures are dropped from the pool).
4. Once etic dimensions are identified, develop emic item scales in each culture that measure the etic construct.

We encourage cross-cultural researchers to develop similar strategies to ensure that their constructs and measures are employed appropriately across samples.

Treatment of Culture

Another important issue in the development of the cross-cultural research question is determining or understanding how culture will be treated in the research design and how it will be operationalized. In 79% of the studies we examined, country was used as a proxy for culture (e.g., Shane, Venkataraman, & MacMillan, 1995; Steensma, Marino, Weaver, & Dickson, 2000). Although country may in fact be a suitable and convenient indicator of culture, using it as the sole operationalization of culture has limitations. The two terms may be incongruent with one another for a number of reasons, including certain national boundaries being set by outside parties, political differences within a country, and a country’s specific cultural identity, which supports more than one subculture (Peterson & Smith, 1997). In many research settings, there may be specific within-country differences along certain dimensions that are greater than between-country differences (Samiee & Jeong, 1994). For example, although some countries such as Japan have relatively homogeneous cultures, other countries such as Canada and Switzerland may have more distinct subcultures within their borders (Peterson & Smith, 1997).

Ryan, McFarland, Baron, and Page (1999) recognized this problem when they examined human resource selection practices in 20 different countries: “A concern is that the use of nation as a basis for examining cultural differences can be criticized as not attending to subcultural differences” (p. 388). The implications of this potential incongruency between country and culture are important. Researchers who inappro-
appropriately use country as a proxy for culture run the risk of not capturing all of the relevant cultural factors that might lend support to (or that might discredit) their theories and hypotheses. Ultimately, the construct of culture should have a theoretical role in the cross-cultural research framework. We discuss this idea further in the Best Practices section below.

We also found that many researchers (44 out of 83 studies) used country as a proxy for Hofstede’s (1980) cultural value dimensions. Rather than measuring respondents’ value orientations directly, researchers often used Hofstede’s ranking of countries, which is now more than 25 years old, to assign value labels to their cross-cultural samples (Earley & Gibson, 1998). For example, C. W. Mueller, Iverson, and Jo (1999) hypothesized that individualistic value orientations in the United States and collectivistic value orientations in South Korea would be related to how distributive justice evaluations are formed in the two societies. Rather than measuring these value orientations directly, the researchers assigned values based on pre-established norms of individualism in the United States and norms of collectivity in South Korea. Similarly, Janssens, Brett, and Smith (1995) relied on Hofstede’s individualism–collectivism rankings when they examined perceptions of worker safety in plant locations in the United States, France, and Argentina. Varying degrees of individualism–collectivism across the three countries were once again assumed, rather than directly measured (see also Greer & Stephens, 2001; Jackson, 2001). Using country as a proxy in these situations can be problematic because, as previously stated, sample differences unique to each research setting might very well be inconsistent with national trends or norms.

An additional issue regarding the treatment of culture is the level of analysis at which relationships are to be observed. One problem in cross-cultural research is that there are often two (or more) levels of theorizing that if not coordinated effectively into a research design, may actually compete with one another. These levels include the individual level, where psychological processes, attitudes, and values are often studied, and the societal level, where political and anthropological trends are common (Hofstede, 1991). An understanding of the different levels within a cross-cultural context is an important prerequisite for analyzing and reporting research results.

For example, if a researcher were interested in value differences across multiple countries, she or he would need to first assess the degree to which values within each sample were similar. Survey data would need to be collected from individuals and then aggregated to the level of country by computing a mean score for each country sample (Hofstede, 1991). Mean scores for country-level values would be meaningful only if there were similarity among respondents from country samples with respect to their individual value rankings. Observed similarity within samples justifies the aggregation of variables from lower to higher levels.

Three methods for dealing with level of analysis issues, which can be utilized in cross-cultural research, are the interrater agreement index ($r_{wg}$), within and between analysis (WABA), and hierarchical linear modeling (HLM). These methods are important in the sense that they can help establish construct validity in research applications that involve multiple levels (Bliese, 2000). $r_{wg}$, an indicator of within-group agreement, provides justification for the use of higher level constructs based on consensus at lower levels (Chan, 1998; James, Demaree, & Wolf, 1984). This index is calculated by comparing the variance of a group’s scores to an expected random variance. Generally speaking, this comparison allows researchers to assess the degree to which
individuals within a selected group (or collective) give the same rating or score on a construct.

WABA is a statistical approach that can be used to examine which of any number of levels may apply for a set of constructs in a study (Dansereau, Alutto, & Yammarino, 1984; Dansereau & Yammarino, 2000). If one were studying cultural values across different contexts, WABA could be used to show within-group similarity for a sample of respondents with respect to value orientations, but it could also reveal the nature of any variability found within the sample. For example, respondents' scores might be interdependent, or they might be independent of one another. Dansereau and Yammarino’s (2000) discussion of WABA focuses on four important issues that can assist cross-cultural researchers in dealing with level of analysis complexities: (a) conceptualizing each level of analysis in a research setting, (b) combining the levels of analysis, (c) associating different formulations of variables with each level of analysis, and (d) identifying conditions where a higher level of analysis (i.e., nation or country) might moderate the development of a lower level (i.e., group or individual).

Although much has been written about comparisons between WABA and rwg (e.g., George & James, 1993; James, 1998; Schriesheim, Cogliser, & Neider, 1998), it is important to realize that the two approaches can be complementary (Schriesheim et al., 1998). For example, Dansereau and Yammarino (2000) suggested that WABA can serve as a preliminary test to ensure that there are group-level differences between groups, whereas rwg can be used to assess the agreement for each group separately. “If one were to find differences between groups (using WABA), rwg could be used to test whether scores within all groups are indeed similar” (p. 444).

The third method, HLM, can also be useful when variables at one level of analysis influence or are influenced by variables at another level of analysis (Hofmann, Griffin, & Gavin, 2000). This is often the case in cross-cultural research when climate or culture and individual behavior are examined together in the same framework. Suppose, for example, a researcher were interested in understanding how individual employee commitment is influenced by individual self-efficacy and individualism/collectivism (measured as an aggregate of responses from each country sample). In this case, the independent variables represent two levels of analysis. One strategy for dealing with this is to disaggregate the data so that lower level units (i.e., each individual) are assigned scores based on the higher level variable. Then ordinary least squares regression could be used to test the hypotheses. Another strategy is to aggregate the lower level variables to a higher level (as discussed above).

Hofmann et al. (2000) suggested that HLM may be the researcher’s best option for these types of cross-level analyses: “HLM explicitly models both the lower-level and the higher-level random-error components, therefore recognizing the partial interdependence of individuals within the same groups (or collectives)” (p. 471). This feature is unique and differs from ordinary least squares approaches that estimate group-level and individual-level random errors separately. For a more thorough review of HLM advantages and procedures, the reader should refer to Hofmann et al. (2000) and Raudenbush and Bryk (2002).

Importantly, not all cross-cultural applications require analyses at multiple or higher levels. For example, Earley (1994) investigated individual-level relationships between cultural value orientations and self-efficacy and performance for Chinese and American respondents. In some cases, cultural variables at the individual level may be
appropriate for predicting individual-level outcomes (Chao, 2000). Nevertheless, con-
siderations of level of analysis and aggregation will likely be important issues for
many cross-cultural research settings, and researchers should familiarize themselves
with the topics that have been briefly presented in this section.

**Best practices.** To ensure the integrity of their cross-cultural research, researchers
should pay attention to whether their treatment of culture is appropriate. We discuss
here two related best practices. First, researchers should minimize the use of country as
a proxy for culture. The specific constructs or variables in a study should be carefully
examined to assess the appropriateness of using other delimiters of culture (besides
country). For example, in using samples from both Australia and Sri Lanka, Niles
(1999) recognized that both settings were multicultural societies and that ethnicity
could be a confounding factor. As such, the samples were drawn not on the basis of
national boundaries alone but also on the basis of ethnic groups within each country
(see also Lenartowicz & Roth, 2001).

Peterson and Smith (1997) provided a comprehensive list of cultural determi-
nants, other than country, that can help researchers with this issue. These determinants
include language, proximity and topography, religion, economic development, tech-
nological development, political boundaries, industry type, and climate. For example,
language differences separate cultural groups because they affect the ease with which
members communicate relationships between symbols and meanings (Peterson &
Smith, 1997). Religious differences also may be a source of cultural variation among
groups of people within the same country because of unique traditions and customs.
These examples highlight only some of the potential factors that can be used in addi-
tion to country to identify sources of cultural differences.

Our second recommendation is for researchers to incorporate culture into their the-
etorical frameworks. This recommendation is consistent with the last one, in the sense
that using country as a proxy for culture can often be viewed as inherently atheoretical.
To date, there seems to be a lack of adequate a priori theorizing on why and how culture
accounts for observed differences. Researchers need to base their designs on contex-
tual variables and theory. For example, Cheng’s (1989) approach to cross-cultural
research focuses on the organization as the primary object of interest and uses key con-
textual variables that vary across nations, including economic, legal, and political
structures (similar to Peterson & Smith, 1997). In a similar vein, Triandis (1994b)
referred to “cultural syndromes,” such as societal complexity, and the tightness/loose-
ness of a culture to illustrate how comparisons across contexts should take into account
such factors. From a theoretical perspective, it may make more sense to ask, How
would employees in tight cultures be expected to differ from employees in loose cul-
tures on a variable such as job satisfaction? rather than, How are employees in Japan
different from employees in the United States on job satisfaction?

In addition, both Erez and Somech (1996) and Aycan, Kanungo, and Sinha (1999)
provide examples of how culture is used theoretically in management research. Erez
and Somech used the distinction between individualism and collectivism (one of
Triandis’s cultural syndromes) to tie culture to another theory, workplace motivation.
For example, individuals with a dominant independent self might find certain work
environments motivating, whereas individuals with a dominating interdependent self
might be motivated by other types of work environments (Erez, 1997). In their culture-
based model of work motivation, Erez and Somech treated culture as a moderator in the sense that what motivates people is influenced by culture. Motivational practices that are inconsistent with employees’ cultural values are suggested to be less effective than congruent practices.

Aycan et al.’s (1999) Model of Cultural Fit uses culture as a critical contingency variable to help explain the use of human resource management practices across a number of different countries. From a theoretical perspective, this research is aimed at addressing the issue of how culture influences organizational processes. Briefly, managers’ perceptions of their own sociocultural environment were examined in relation to their assumptions about employees and human resource management practices in their organizations. Data on these relationships were analyzed for each country to determine whether culture would act as a moderator of the proposed relationships (Aycan, 2000; Aycan et al., 2000).

These studies illustrate how culture can be effectively incorporated into a theoretical framework. Such approaches differ sharply from other cross-cultural works that have explained cultural influences in a post hoc, exploratory fashion (Aycan et al., 1999). In the preceding examples, culture was treated as a moderator, but other theoretical applications are also possible. We refer the reader to Brett, Tinsley, Janssens, Barsness, and Lytle (1997), who discuss the use of culture theoretically as a main effect, as a moderator, or as an influence on the meaning of constructs.

Finally, related to the important role of theory in cross-cultural research, researchers should take care to reflect on the nature of cross-cultural differences found in their studies (Van de Vijver & Leung, 2000). Some cultural differences are ingrained in society, whereas others seem to be more narrowly applicable to the particular research setting. This is an important distinction in that the former type of difference would be more generalizable to other processes or constructs beyond the scope of the focal study, with the latter type being more limited in scope (but not necessarily less meaningful). For instance, observed differences in individualism and collectivism might represent deeply rooted cultural differences. If observed variations in pay preference, for example, were based on such value differences, then you might expect other processes to be similarly affected (e.g., preferences concerning group work vs. individual work). On the other hand, more superficial differences, such as handshaking styles and promptness for meetings, might be relevant cross-cultural differences but more limited in terms of their use for exploring other individual, group, or organizational phenomena. We recommend that these types of considerations be included when researchers interpret their findings.

The final best practice recommendation for this section concerns Hofstede’s cultural value dimensions. We urge researchers to directly measure these dimensions in the specific research context. Adhering to this practice can be difficult, however, because the conceptualization of these constructs has been inconsistent across different studies. As a result, different measures for these dimensions have been used and discarded over the years (Earley & Gibson, 1998). For example, numerous measures of individualism-collectivism have appeared in the literature, and a consensus has not been reached as to which one is the best. This construct was measured in a variety of ways in our sample of studies: Birnbaum-More, Wong, and Olve (1995) measured individualism-collectivism with Hofstede’s (1980) Values Survey Module; Tinsley (2001) used a scale developed by Earley (1993); Casimir and Keats (1996)
used the INDCOL scale (Bontempo, 1993); Jung and Avolio (1999) used a scale developed by Bass and Avolio (1997); Thomas (1999) used an 8-item subscale from Maznevski, DiStefano, Gomez, Nooderhaven, and Wu (1997); and Smith, Dugan, and Trompenaars (1996) derived their own survey items to measure individualism-collectivism.

This inconsistency of measures across studies presents a challenge for researchers and reflects the difficulty involved in specifying the items needed to reveal important value dimensions. Despite this challenge, we still believe efforts to operationalize these dimensions in the particular research setting will be more fruitful than relying on pre-established categorizations based on Hofstede’s (1980) country rankings or numeric ratings. One recommendation is for researchers to use multiple methods to identify cultural differences (Schwartz, 1994b; Triandis, McCusker, & Hui, 1990). Observing convergent validity among more than one measure would provide more confidence that the chosen measures are effectively capturing targeted cultural values.

Schwartz’s (1994a) values study illustrates a conceptual and operational approach for deriving cultural values. Representing an alternative to Hofstede’s classification, this study developed a survey that measured the content of individual values across cultures. Ten broader types of values (power, achievement, hedonism, stimulation, self-direction, universalism, benevolence, tradition, conformity, and security), represented by 56 specific values, were validated in this study. Schwartz, using Smallest Space Analysis, was able to evaluate the match between observed and theorized structures of value types (Schwartz, 1994a). For the most part, his analyses confirmed all a priori hypotheses concerning the dimensionality and content of the values. These findings supported the idea that different respondents across more than 40 countries would be able to discriminate all 10 value types when asked to rate the importance of their values (Schwartz, 1994a). We are not suggesting that Schwartz’s value dimensions are necessarily the only ones to be used when conducting cross-cultural studies. However, his procedures for developing measures of values is an example of how one can avoid relying on more rigid categorizations, such as Hofstede’s rankings.

Summary

The development of a cross-cultural research question involves at least three important issues. First, the researcher must consider the difficulties involved in balancing the etic and emic approaches. Cross-cultural comparability is an appropriate goal for researchers as long as they consider the idiosyncratic aspects of each particular culture in the study. In this section, we discussed the benefits of using a combined etic-emic approach in which emic dimensions are first generated for each culture and then analyzed alongside each other to determine where etic comparisons would be appropriate. Second, the researcher must be aware of the potential difficulties involved in using country as a proxy for culture or as a proxy for Hofstede’s values. If the cross-cultural samples in the study come from countries with relatively homogeneous populations, then this issue might not be as pressing. On the other hand, when the countries have heterogeneous populations, researchers need to be aware of other cultural determinants and must recognize that within-country differences may be inconsistent with pre-established national categorizations. Finally, researchers must be cognizant of incorporating culture into theoretical frameworks within their research designs.
Stage 2: Alignment of Research Contexts

Several aspects of the research contexts must be considered when designing cross-cultural studies. The alignment of contexts refers to establishing congruence between the different cultures being studied. Without appropriate contextualization, researchers may interject ethnocentric attitudes and perspectives into their study designs (Chikudate, 1997). These attitudes and perspectives may conceal important cultural differences between the home-base culture and the other cultures that will be compared to it. In our review, two main issues were particularly relevant to this stage. Researchers should address the equivalency of their samples across contexts as well as the uniformity of their survey-administration procedures.

Equivalence of Samples

An important issue with respect to contextual alignment is whether cross-cultural samples are equivalent on dimensions other than the ones under examination. Researchers need to minimize the effects of sample differences that are not relevant to the main purposes of their studies. Paying attention to such differences may be particularly important in organizational cross-cultural research. For example, certain demographic imbalances within organizations may be more prevalent in some cultures than in others, employees in some samples might be expatriates whereas others might be host-country natives, organizations might be from different industries, and employees might have differing levels of experience. To the extent that such factors are taken into consideration, and controlled for, cross-cultural researchers should have more confidence that any differences detected are due to hypothesized cultural differences rather than to these other types of factors.

Demographic differences across samples were noted in Aycan et al.’s (1999) study that compared Indian and Canadian employees on measures testing the Model of Culture Fit (see Kanungo & Jaeger, 1990; Mendonca & Kanungo, 1994). Results of the study showed, among other things, that Indian respondents scored higher than Canadian respondents on paternalism, power distance, uncertainty avoidance, loyalty to community, reactivity, and futuristic orientation. However, the Indian sample was significantly older and had a comparatively higher average education level than the Canadian sample, and the Canadian sample was much more balanced in terms of gender proportions (Aycan et al., 1999). Differences such as these are important when they exist alongside cultural differences because they can affect a study’s results. For instance, might the higher ages of the Indian respondents in this example be related to their higher scores on paternalism and loyalty to community? Importantly, in this particular study, the researchers used statistical analyses to covary these demographic differences out. This is a best-practice approach that we discuss in a later section.

Environmental characteristics can also be a concern in cross-cultural research. Surrounding organizational and social environments can be problematic to the extent that they vary across samples. For example, some employees working in multinational subsidiaries are natives of the host country, whereas others are working as expatriates. Individuals working in a foreign country are exposed to a much different environment than that of individuals working in their own native country, and a concern in some cross-cultural research might be that such differences in work environments could contribute to differences in respondents’ values, preferences, and attitude (Beldona,
Inkpen, & Phatak, 1998). Environmental factors that can differ across samples represent those types of cultural delimiters that we have already discussed (e.g., topography, religion, economic development, technological development, political boundaries, and climate). Therefore, in addition to micro-level factors such as demographics, researchers should consider these types of macro-level characteristics when thinking about sample equivalence.

Sampling across different industries might also present challenges. In mailing out surveys to 300 organizations in 22 countries, Ryan, McFarland, et al. (1999) used a random sampling strategy in selecting firms that was based on a minimum number of employees in each organization. The researchers examined differences in selection practices across different cultures. An important issue in this study is the extent to which industry differences within the sample might have contributed to the variability in selection practices across the different firms. In fact, the researchers themselves suggested that other sampling strategies, which take into account industry characteristics, might have allowed for better comparisons across countries (Ryan, McFarland, et al., 1999).

Differences in respondents’ job-related experience can be a problem for researchers when they are trying to maintain sample equivalency. This issue was present in Merritt and Helmreich’s (1996) study that examined the influence of culture on pilots’ and flight attendants’ attitudes toward group processes and performance levels on the flight deck. Surveys were administered to pilots from the United States, the Philippines, and Taiwan and to flight attendants from the United States, Hong Kong, Japan, Korea, and Thailand. Results of the study indicated some key differences between the U.S. and Asian respondents. For example, all flight attendant groups in the low power-distance U.S. culture preferred a captain who encouraged questions and participation, whereas most of the flight attendants in the high power-distance Asian culture preferred captains who had a more autocratic style (Merritt & Helmreich, 1996). Although this finding has cross-cultural implications, other sample differences may have had an effect on the results. Specifically, the U.S. flight attendants had an average of 10 years or more of flying experience, whereas the Asian flight attendants had an average of only 2 to 3 years of experience. Perhaps the flight attendants’ attitudes toward leadership styles in their captains were as much a function of this difference in experience as they were a function of cultural differences. In other words, a flight attendant with only 2 years of experience might prefer a captain with a more direct and authoritative style, and a flight attendant with 10 years of experience (someone who knows the ropes and has suggestions to offer based on past experiences) might prefer a captain who expects contributions and ideas from others.

Finally, cross-cultural samples can also differ in terms of the experiences respondents have had with measurement instruments and with general testing procedures. There may be important differences across cultures in terms of the familiarity people have with filling out surveys. For example, compared to other cultures, most Western citizens are relatively familiar with test and survey formats and are more comfortable in completing measurement instruments (Lonner, 1990). This issue can be particularly important in cross-cultural studies that examine constructs across many different countries because researchers might have more trouble assessing differentials of familiarity across the culturally diverse groups of respondents (e.g., see Arthur & Bennett, 1995; Geletkanycz, 1997). In such studies, greater variation in respondents’
comfort levels and/or prior exposure to the testing formats can influence item responses and may have implications for the interpretation of obtained results.

Best practices. As noted above, efforts should be made to match samples in terms of many characteristics so that sample differences can be ruled out as alternative explanations for results thought to be due to cultural differences (Van de Vijver & Leung, 1997).

Among the cross-cultural studies we reviewed, 56% used this best practice of matching samples. Tinsley (1998) used a matching strategy in her study of conflict resolution styles among managers from Germany, Japan, and the United States. In each of the three cross-cultural samples, participants had been educated by business programs in their culture and were similar in terms of age and gender. Tinsley noted that “matching participants enables noncultural characteristics to be ruled out as alternative explanations for observed differences in conflict model usage” (p. 319). In a similar fashion, Giacobbe-Miller, Miller, and Victorov (1998) matched U.S. managerial respondents to Russian respondents on industry background, age, and gender (see also Begley & Tan, 2001; S. L. Mueller & Clarke, 1998; Pavett & Morris, 1995). In each of these examples, the researchers proactively matched samples along important characteristics, thereby increasing overall levels of equivalency.

While advocating the use of matching samples as a best practice in cross-cultural research, we also provide a word of caution for researchers. A potential problem in matching samples is that when matching on one set of variables, researchers may at the same time be matching on a related cultural variable, thus restricting samples and masking cultural differences. Consider a situation wherein researchers match samples of top executives across cultures based on gender. If the samples contain a consistent mix of women and men across cultures (say 50-50), the researchers would need to consider how the cultures might differ in terms of how easily women in the general population ascend to top management positions. For instance, there may be key differences between some female top executives in the United States and matched female top executives in Eastern cultures that could be due to societal/cultural differences. This example highlights the need for researchers to take special care when matching samples. Matching should not be done as a blanket practice across all studies and all situations but should be done only after researchers have thoroughly considered the link between these variables and other cultural factors.

We also provide a word of caution related to the common practice of matching samples by using college students as respondents (e.g., S. L. Mueller & Clarke, 1998). Students across different cultures are often used because it is generally assumed that they are similar along a number of characteristics, especially demographics (Van de Vijver & Leung, 1997). In some cases, because of similarities in age and educational groupings, there may also be strong similarities in students’ attitudes, values, and belief systems. Researchers should be aware that these types of similarities could potentially mask certain cultural differences that would have otherwise been observed if nonstudent samples had been used. On the other hand, student samples may differ in significant ways, such as being enrolled in different subjects, having different college majors, and coming from different types of universities (Strohschneider & Guess, 1999; Watkins et al., 1998). These types of differences can contribute to variations in outcome measures. Therefore, when matching strategies involve college students, researchers
should be aware of both potential similarities and potential differences across samples that would be inconsistent with their underlying theory. Other factors beyond the classification of student/nonstudent can contribute to sample equivalence (or sample differences).

Finally, when matching samples, we recommend that researchers assess the degree to which there might be differences in comfort levels with filling out surveys. In many cases, this is more difficult than assessing objective characteristics, such as demographics. Researchers may benefit by incorporating this type of evaluation into earlier research stages, where emic analyses are typically conducted. This would allow researchers to discover initial incongruities in this area and identify samples that might need training or some type of orientation session.

Researchers will often find it difficult to use a matching strategy because resources and subjects have varying degrees of availability and because different cultural groups often have contrasting profiles along important characteristics (Hudson, Baraket, & LaForge, 1959; Sekaran & Martin, 1982; Van de Vijver & Leung, 1997). Therefore, it may be necessary to statistically control for the differences that remain between the samples. In our sample, 37% of the studies used this best-practice procedure to deal with sample differences (e.g., Birnbaum-More et al., 1995; Greer & Stephens, 2001; Steiner & Gilliland, 1996).

For example, when Peterson et al. (1995) compared managers’ perceptions of role overload and role conflict in 21 different nations, they realized that an initial strategy of matching samples based on industry would not account for cultural differences. As a result, they selected a wide array of demographic and organizational characteristics (age, gender, organizational size, task, years of education, and departmental experience) to use as statistical controls. Similarly, when Buda and Elsayed-Elkhoul (1998) examined cultural differences between Arabs and Americans, they had to statistically control for age, gender, education, industry, and management level. Analysis of covariance was used to examine differences on individualism-collectivism scores, and the results showed a significant main effect after the influence of the control variables had been removed (Buda & Elsayed-Elkhoul, 1998).

As with the matching of samples, researchers should take care when controlling for variables. If the variables are linked in any way to cultural or societal differences, then caution should be exercised when partialing out that variable’s effect on important outcome variables.

**Administration of Surveys**

Another important issue related to contextual alignment is whether the administration of surveys is consistent across different research settings. In this section, we discuss the need for cross-cultural researchers to establish equivalence in their data-collection procedures and to maintain consistent levels of rapport with respondents from different cultural backgrounds.

Procedurally, there should be consistency across samples in terms of survey formats, data collection, and survey timing. For example, if items of an instrument were read to a sample of respondents in one culture (because of low literacy levels), and administered in written format in another culture, the measurement reliability and validity of the study could be compromised due to this administration inconsistency (Ortega & Richey, 1998). This practice was observed in Rahim and Magner’s (1995)
study of conflict-handling styles in the United States and Bangladesh. The U.S.
respondents filled out a written survey, whereas the Bangladesh respondents were
interviewed and thus gave answers orally to the same set of items. Aulakh, Kotabe, and
Teegen’s (2000) study of firms’ export strategies in Brazil, Chile, and Mexico also
illustrates a case of procedural inconsistency. The data collection in this study varied
across contexts because of certain limitations or opportunities inherent in each coun-
try. The procedures, which included mailings, hand-deliveries, faxes, and student con-
tacts, were applied unequally across the three samples.

The timing of survey instruments in cross-cultural research is another important
part of establishing consistency across contexts. For comparison purposes, data should
be collected in all cultures simultaneously or at least within a reasonable time period
(Yu, Keown, & Jacobs, 1993; Sekaran & Martin, 1982). Too often, cross-cultural
researchers combine survey responses that were collected at different points in time,
giving little attention to whether seemingly similar conditions across the samples were
really the same (Roberts & Boyacigiller, 1984; Roberts, Hulin, & Rousseau, 1978).

Korean teachers exemplifies how timing issues can be problematic with respect to sur-
vey administration. The researchers used a sample of U.S. teachers in Chicago and a
sample of South Korean teachers in Seoul. Survey instruments were administered to
the U.S. sample between the years of 1986 and 1990 and to the South Korean sample in
1994. This difference in survey timing may have potentially allowed other factors, in
addition to cultural differences, to contribute to the variability in responses across the
samples. An example would be certain historical effects associated with global or soci-
etal changes during the 9-year period between 1986 and 1994. Although we are not
suggesting that such factors necessarily came into play in C. W. Mueller et al.’s study,
we nevertheless use it as an example to illustrate the potential difficulties that can
result from this aspect of procedural inconsistency in survey administration (for a sim-
ilar example, see Elenkov, 1997).

As an additional step in establishing procedural equivalence, researchers should
make efforts to maintain similar levels of rapport with respondents. Rapport in this
sense refers to the respondents’ confidence in the researcher, their overall comfort
level with the researcher, and /or their willingness to cooperate with the procedures
associated with the survey instrument. We are not saying here that researchers should
necessarily establish higher (or unequal) levels of rapport with foreign respondents.
However, we do feel that rapport is a factor within the research environment that
should be considered when researchers try to establish overall equivalence.

Rapport problems can arise when respondents feel that the researchers are being
obtrusive, especially when the two groups come from culturally different backgrounds
(Van de Vijver & Leung, 1997). If respondents feel less comfortable with the research-
ers because of such differences and if they hurry through the instrument to decrease
this level of discomfort, their response patterns to survey questions may be affected
(Anastasi, 1988). Also, there may be differences in respondents’ perceived levels of
status discrepancy between themselves and the researcher (Yu et al., 1993), and in some
cultures, respondents may be more apprehensive in the presence of an outside
researcher. Both of these conditions could influence respondents to conform to what
they think the most popular answer would be (see K. Yang, 1981). Many cross-cultural
research studies are subject to these types of rapport problems, especially if one
researcher is administering the survey instrument across the different samples.
**Best practices.** With respect to survey administration, we recommend that cross-cultural researchers make efforts to establish consistency across samples in terms of data collection and instrument formats. In addition, explicit instructions and examples should be included in all survey instruments, and these should be provided to each of the samples in a consistent manner.

Survey timing should be a consideration when initial theories and hypotheses are specified. In addition to controlling for variables such as industry type and organizational size, cross-cultural researchers should also consider time-related factors, such as organizational life cycles (Roberts & Boyacigiller, 1984). This may be particularly important if researchers are examining longitudinal organizational processes. Business cycles are often influenced by external environments and societal norms. For example, Van de Ven and Poole (1995) noted that certain stages of development in U.S. organizations can be influenced by the “natural order of Western business practices” and by institutional rules and policies. Across different cultures, such factors may vary and should therefore be considered possible influences on proposed relationships.

When researchers from our sample of studies mentioned procedures related to the equivalence of their survey administrations, most often they made statements describing consistent instruction formats or consistent survey-distribution procedures. Both Aycan et al. (1999) and Casimir and Keats (1996) explicitly described the procedures they used to make sure all of their respondents got the same detailed instructions and explanations related to the survey process. In addition, a handful of studies provided specific information about procedures related to data collection protocols, mailing procedures, and/or the personnel responsible for administering the surveys (see Pulich, Hom, & Griffeth, 1995; Salk & Brannen, 2000; Song, Di Benedetto, & Zhao, 1999; Yousef, 2000).

Despite these examples, we found it hard to identify a specific study exhibiting most or all of the aspects related to equivalent survey administration procedures. We do not necessarily interpret this as researchers ignoring important technical aspects of the survey process. Rather, we suspect that in most cases, certain survey-related procedures have just not been reported. Our view here is that such procedures should be reported, in the sense that as cross-cultural research continues to grow, such procedures described in studies can serve as benchmarks for other researchers who are looking to engage in similar types of endeavors.

Finally, we suggest that researchers attempt to maintain uniform levels of rapport across samples so that both respondents from foreign cultures and native respondents feel equally comfortable with the intervention. We recognize, however, that this objective can be problematic in cross-cultural research because, in some cases, it can actually be counter to the goal of standardization. For example, a U.S. researcher might spend some time with non-U.S. respondents prior to the administration of a survey to give them a chance to express concerns or ask questions about the instrument and the procedures. This type of interaction is typically discouraged in cross-cultural research because the researcher in this situation can introduce nonstandardized and potentially important sources of variation (Van de Vijver, 1993). A difficult balance is therefore required. The researcher should attempt to make respondents feel comfortable in the research setting, but the interactions between the researcher and respondents should be standardized across samples, based on explicitly described roles that were set up prior
to the intervention. Working with local collaborators is perhaps one way to do this. Under this scenario, the host nation researcher can collaborate with foreign researchers to establish uniform procedures. This should decrease respondents’ perceptions of the researcher’s being obtrusive and should increase equivalence across research settings in terms of rapport levels. We found a small number of studies (8%) that mentioned issues related to levels of rapport. For example, Kozan and Ergin (1998) implemented procedures to ensure that the research assistants were introduced by name to respondents in both Turkey and the United States. This effort helped establish a level of rapport in both locations (see also Song et al., 1999; Teagarden et al., 1995).

In summary, there are a number of issues that come into play when cross-cultural researchers attempt to align the research contexts. This section has presented the ones we feel might represent the biggest challenges. Researchers should be aware, however, that other factors might also contribute to variability across contexts. Such factors include the use of ethical guidelines, the procedure of random assignment, and differential perceptions of demand characteristics. Triandis (1992, 1994a) discussed the importance of the ethical acceptability of the research methodology. In cross-cultural research, it is important to establish standards that apply to all samples consistently. For example, respondents from different cultures should be provided the same information regarding the research project, and they should be given the same opportunities to participate in (or withdraw from) the study. In addition, researchers should be similarly familiar with all of the cultures in the study.

To do research that is ignorant of or insensitive to the major features of the local culture often means to do poor research and thus wastes the time of local subjects, as well as the funds, and that is unethical. (Triandis, 1992, p. 232)

One solution that is offered is for researchers to establish teams that collaborate with one another, with each collaborator being an expert on one of the cultures in the study.

With respect to the fairness of random assignment, it should be noted that assignment to a research project could be either a positive or a negative experience for an individual. Assignment procedures can be differentially fair to people (Baier, 1985; Griffin, 1985), and they can cause ill feelings toward the research or researchers in certain cases. Individuals in different cultures might have varying degrees of feelings about being assigned to the research project, and these different feelings could set the table for some differences in survey responses. Another factor, susceptibility to demand characteristics, can also vary across cultural settings. Respondents may have differential access to cues that might convey the purposes of the research. These cues include rumors about the research, information conveyed during the orientations, the actual researchers themselves, the setting in which respondents fill out surveys, and any other implicit or explicit communications throughout the research process (Orne, 1962). These types of demand characteristics, if not accounted for, can vary across samples from different cultures.

These additional factors are illustrated briefly here to highlight the need for researchers to continually monitor the contexts in which they conduct their studies. As we suggested at the beginning of this section, ignoring potential sources of misalignment can conceal important cultural differences between samples.
Stage 3: Validation of the Research Instruments

An important problem in cross-cultural research is that many studies examine common organizational topics across cultures without fully taking into account critical measurement issues (Erez, 1994; Ryan, Chan, Ployhart, & Slade, 1999; Triandis, 1994a). Researchers must ensure that the measures of a construct developed in one culture can be applied to another culture before they can establish a basis for theoretical comparisons. In this stage, we review three types of equivalence related to this issue. First, linguistic differences among cross-cultural samples may affect the semantic equivalence of multiple versions of a research instrument. The task of translating instruments across different languages often presents problems for cross-cultural researchers (Holtzman, 1968). Second, researchers need to ensure that there is conceptual equivalence of measures across the samples. Related to this issue is whether the survey instruments elicit the same conceptual frames of references in culturally diverse groups of respondents. Finally, scaling equivalence concerns the degree to which scoring formats on instruments are interpreted or calibrated in the same way across samples.

Semantic Equivalence

In establishing semantic equivalence, the researcher’s goal should be to ensure that multiple versions of a self-report instrument used cross-culturally fully account for linguistic differences among the samples. The meaning of each item after translation should be consistent for the different respondents from each culture. This is rarely an easy task. Even in situations where researchers and linguists work together to produce a common version of an instrument, it is still possible that remaining underlying differences in meaning will present threats to the interpretation of findings (Holtzman, 1968).

The following survey items from our sample of studies show how certain phrases can be problematic for translation purposes: “Management here does not cut corners where safety is concerned” (Janssens et al., 1995), “Major bottlenecks prevent improved operations” (Gunther-McGrath, MacMillan, & Venkataraman, 1995), “I automatically tune myself in to other people’s expectations of me” (Gabrielidis et al., 1997), and “I think that wanting to be a company man or company woman is sensible” (Yousef, 2000). Such items are everyday expressions in the United States but may not survive the translation process very well (Small et al., 1999). Careful revision during translation may be necessary for them to convey the same general meaning across cultures.

In our review, researchers often mentioned the use of back-translation (a best practice discussed later) but made no further comments about issues related to semantic equivalence. This is important because even with back-translation, semantic inconsistencies may still remain. In translating a scale into Vietnamese, Small et al. (1999) found that one item on the instrument (“I have been looking forward to things with enjoyment”) became partially distorted after the back-translation process. The back-translation resulted in two phrases that were not consistent with the original English version (these were, “I have been hoping/expecting to be happy” and “I have been feeling optimistic”). This example shows that even with careful and thorough translations,
survey items may still contain peculiarities in meaning that affect the ultimate results of the study.

Finally, some researchers have recognized that certain words on measurement instruments can subtlety elicit either cognitive or emotional states and that this distinction can affect semantic equivalence during the translation process. For example, individuals might respond differently to survey items that are phrased cognitively versus affectively (Ortega & Richey, 1998; Ponterotto & Casas, 1991). The two statements, “What do you think about your supervisor?” and “How do you feel about your supervisor?” are very close in meaning but are categorically discernible based on their respective cognitive and affective orientations. The important issue, which has still not been explored fully in organizational research (Ortega & Richey, 1998), is whether these orientations will become ambiguous during translation processes. In other words, will an affective item still elicit an affective response after translation, and will a cognitive item still elicit a cognitive response?

**Best practices.** For establishing semantic equivalence, we recommend that researchers employ back-translation before administering an instrument to respondents who speak a different language from the one in which the instrument was originally developed and validated. In this process, bilingual experts translate the instrument from Language A to Language B and then back again to Language A (Ortega & Richey, 1998). The purpose of this double translation is to allow experts to examine each survey item on both versions to establish meaning conformity. If inconsistencies are found, items can be reworded or, if necessary, eliminated. Often, translators in this process have had some experience with the organization or the setting in which the research is going to take place (e.g., Ryan, Chan, et al., 1999). This is advantageous because they may have some inside perspectives that would better qualify them to identify specific phrases or idiosyncrasies that would uniquely apply to the organization. Janssens et al. (1995) used back-translation when they took an English version of a workplace safety questionnaire, translated into French and Spanish, and then back into English to ensure the similarity of meaning across the three versions (see also Peterson et al., 1995). In our sample of studies, 62% indicated that this process was used.

In many research settings, respondents across samples from different cultures or countries share a common language (e.g., English). An important issue in this type of setting is whether researchers should translate the surveys into respondents’ native languages or administer the English version to all respondents. Importantly, there are trade-offs involved in such a decision. Administering the survey across the samples in identical language versions might provide more item equivalence, but it might affect response patterns in important ways. For example, research has shown that non-U.S. respondents might give more extreme answers in their responses when they complete surveys in English versus when they respond in their native language (Bennett, 1977; Triandis, 1994a). In addition, people may better reflect their cultural values and assumptions when they respond in their native language. The implications of such tendencies should be considered when researchers are faced with a translation–no translation decision. If there are particularly difficult problems anticipated in translation, administering English versions might provide the best option for preserving item equivalence. However, if the response biases mentioned above are a concern, then it
may be in the researcher’s best interest to work through the translation problems and administer the surveys in the respondents’ corresponding language versions.

For ensuring semantic equivalence, we also recommend that researchers avoid using certain figures of speech, terminologies, or phrases in their survey instruments that may be common in the home-base culture but unfamiliar to other cultures. Related to the above discussion, this may be particularly important when the second culture is English speaking and is responding to an English version of the survey. For example, consider the phrase “I put everything I have into my work.” Respondents from non-U.S. cultures may interpret this saying in a number of different ways. Does the phrase refer to how much effort you put forth while doing your job, or does it mean taking all of your possessions and applying them to the work you do? This same kind of ambiguity can also exist for translators when they are attempting to convert such items into another language. This is especially true when the translator is a native of the foreign culture. In Geletkanycz’s (1997) study, this best practice was demonstrated in the development of an instrument that was ultimately administered to French, German, Italian, Japanese, Spanish, and English respondents. The survey was “designed with an emphasis on common business terminology and minimal use of idiomatic statements” (Geletkanycz, 1997, p. 622; see also Greer & Stephens, 2001).

Two additional best practices that can help with semantic equivalence are (a) using cross-cultural survey instruments in pilot studies and (b) considering insiders’ and outsiders’ perspectives together when developing the instruments. Pilot studies can be used to test the consistency of an instrument across samples and can also be useful for using and comparing several response procedures for measuring the same construct (Van de Vijver & Leung, 1997). Issues identified by researchers using pilot studies are typically related to translation problems and specific ambiguities associated with item phraseology. Therefore, pilot studies can serve the purpose of discovering methodological problems in cross-cultural studies that would have otherwise affected the validity and reliability of results. Researchers can also use information from pilot studies to help them actually develop the measures that will be used cross-culturally.

Hitt, Dacin, Tyler, and Park (1997) used pilot testing when they studied the strategic orientations of U.S. and Korean executives. Prior to administering their survey in the focal study, they conducted a preliminary test where executives from several countries were given the questionnaire. Overall, the pilot study showed that the survey questions were viable and that they were appropriate for measuring the constructs of interest in the main study. In a similar fashion, Aycan et al. (1999) used pilot testing in both Canada and India as a precursor to the development of their questionnaire, which was designed to examine the model of cultural fit.

Obtaining both insiders’ and outsiders’ perspectives together can also help identify some problematic issues. Erkut, Alarcon, Coll, Tropp, and Garcia (1999) described a “dual-focus” approach to developing and validating cross-cultural instruments wherein researchers or “experts” from the indigenous culture become full and equal members of the research team. For example, if an instrument were going to be applied to two samples from cultural backgrounds where language differences exist, the intent would be to have bilingual and bicultural researchers from each culture working together in a team environment. In addition, monolingual members of the unfamiliar culture should be included as members of the team because the language from the other culture (Erkut et al., 1999) would not influence their speech patterns. Although bilingual members of the research team focus primarily on issues related to nonequiv-
alence, monolinguals’ perspectives are important because they spend more time “examining whether the language is stilted (or awkward) or whether it sounds natural to the ears of the native speakers” (Erkut et al., 1999, p. 212). Johns and Xie (1998) used insiders’ and outsiders’ perspectives together when they examined how perceptions of absenteeism would compare across samples from Canada and the People’s Republic of China. The researchers worked together with top executives from a Chinese firm to clear up ambiguities about questionnaire items.

In summary, some important best practices for establishing semantic equivalence in cross-cultural research are to use back-translation, to avoid the use of common figures of speech in survey items, to be cognizant of words or phrases that elicit cognitive or affective states, to use survey items in pilot studies, and to consider both insiders’ and outsiders’ perspectives together. In addition to these best practices, we briefly mention here some suggestions that are specifically geared toward the researcher who is writing a new instrument for a cross-cultural study. Brislin (1986) offered a set of guidelines for optimizing the transferability of items in newly written instruments. From these guidelines, we suggest that researchers use short, simple sentences (less than 16 words), repeat nouns instead of using pronouns (because pronouns may have vague references), and add sentences to provide context for important ideas.

Finally, cross-cultural researchers need to explicitly describe the procedures they used to establish semantic equivalence. Most of the studies in our review (58%) did not include statements about semantic equivalence. For cross-cultural studies to be properly evaluated and replicated, these kinds of statements become necessities.

**Conceptual and Scaling Equivalence**

In validating the survey instrument, researchers must also be concerned with whether survey items elicit the same conceptual frames of references across different cultures (conceptual equivalence) and whether respondents perceive and interpret rating-scale intervals in the same manner (scaling equivalence) (Riordan & Vandenberg, 1994). Often, surveys are routinely administered in cross-cultural research without addressing these concerns. Only 25% of the studies in our sample described procedures related to these types of equivalence. Measurement instruments lacking such equivalence can lead to inaccurate conclusions about important relationships as well as to misguided interventions (Ryan, Chan, et al., 1999).

In cross-cultural research, constructs and their meanings should apply equally across the different cultures being studied. In this vein, conceptual equivalence refers to the degree to which members of different cultures use a common frame of reference when responding to items on a survey instrument (Riordan & Vandenberg, 1994). For example, a group of Japanese respondents might think about the word *bicycle* differently than a group of U.S. respondents (Yu et al., 1993). The former might view a bicycle as a form of transportation, and the latter might view it as a form of recreation. If both groups were administered a survey with items asking questions about a bicycle, the resulting data would likely have equivalence problems. A common construct in organizational research, turnover, may be subject to the same type of problems. Groups of employees from different cultures may have different preconceived notions about the acceptability of turnover in their work environment. For example, one group might view turnover as being more or less taboo, because of societal norms dictating loyalty to the organization, whereas another group (from a different culture) might
view it as being an acceptable part of one’s career development process. In this sense, survey items meant to assess both groups’ attitudes toward turnover could contain underlying inconsistencies related to varying conceptualizations. Researchers should be aware that comparisons between cross-cultural samples should be made only after these types of equivalence problems have been dealt with (Riordan & Vandenberg, 1994).

Survey instruments in cross-cultural research are often in the form of Likert-type scales or semantic differential scales. Importantly, respondents across different cultures do not always interpret or calibrate the scoring formats on these instruments consistently. For example, on a 5-point Likert-type scale, a middle response might mean no opinion to a group of American employees, and the same response might mean mild agreement to a group of Korean employees (Riordan & Vandenberg, 1994). This issue addresses scaling (or true-score) equivalence and presents yet another concern for cross-cultural researchers.

A problem related to scaling equivalence in cross-cultural research is the tendency for certain cultural groups to differ in their response sets (Triandis, 1994a). For example, studies that have compared individualistic and collectivistic samples have shown that there is a greater tendency for collectivists to use the middle or undecided category on a Likert-type scale. Such findings should alert researchers to potential hazards in their own cross-cultural settings. Differences in response sets across samples may very well be due to inherent cultural differences. In Asian cultures, for example, a tendency to respond neutrally to survey items might be related to cultural norms dictating modesty and cautious responses, whereas in Western cultures, a tendency to use extreme responses might be related to norms that endorse individual expressiveness (Triandis, 1992).

Cultural differences might also affect acquiescence response styles, in which respondents have the tendency to agree with a statement regardless of its content. Some cultural groups may be more prone to agreeing with survey items even if they run counter to their true feelings (Javeline, 1999). Differences in this type of response bias across cultural groups may again be due to societal norms, which in this case would dictate how positive or agreeable one should be when responding to survey items. If not addressed by researchers, these types of response biases might mask respondents’ true attitudes or feelings toward the construct(s) of interest, thus creating unwanted sources of variation (Heide & Gronhaug, 1992).

Dealing with response sets in cross-cultural settings can be difficult. Triandis (1992, 1994a) suggested that with large and heterogeneous surveys, one can standardize the data within cultures prior to making cross-cultural comparisons. By converting respondents’ scores to z scores, a researcher can essentially eliminate response sets, allowing for more accurate comparisons (Triandis, 1994a; see also Leung & Bond, 1989). Ohbuchi, Fukushima, and Tedeschi (1999) and Gelfand and Realo (1999) both used this procedure. However, without large and heterogeneous questionnaires, it may be inappropriate to standardize scores, and so we caution researchers against using this as a blanket practice for all cross-cultural studies.

Best practices. For assessing conceptual and scaling equivalence, we recommend two best-practice statistical approaches that have been previously established by researchers. These are (a) covariance structure analysis (e.g., Cheung & Rensvold, 1999; Riordan & Vandenberg, 1994; Ryan, Chan, et al., 1999; N. Yang, Chen, Choi, &
Zou, 2000) and (b) item response theory (e.g., Butcher & Han, 1996; Ellis, Becker, & Kimmel, 1993; Hambleton & Kanjee, 1995; Hulin & Mayer, 1985; Ryan, Horvath, Ployhart, Schmitt, & Slade, 2000).

Typically, researchers have used covariance structure analysis to directly test their equivalency assumptions by placing a series of nested constraints on selected parameters across the samples under study (Riordan & Vandenberg, 1994; Ryan, Chan, et al., 1999; Vandenberg & Lance, 2000). Under this approach, both conceptual and scaling equivalence can be examined in a series of increasingly restrictive hypothesis tests. Cross-cultural researchers can determine equivalence by observing the same number of constructs and items loading on a factor, along with an invariance of factor loadings (Ryan, Chan, et al., 1999). Importantly, these approaches to examining equivalence allow the researchers to specify constraints a priori, with some theoretical justification for proceeding with the analyses (Vandenberg & Lance, 2000).

Riordan and Vandenberg’s (1994) examination of three work-related measurement instruments across samples of Korean and American employees highlight three phases of this covariance structure analytic approach. The first phase involves testing the equality of the variance-covariance matrices across the cultural groups. A rejection of the null hypothesis in this test indicates that the underlying constructs being measured differ across groups in some way. This suggests that increasingly restrictive tests should be conducted to identify the source of the difference (Byrne, 1989; Riordan & Vandenberg, 1994). The second phase specifically assesses conceptual equivalence. Differences in factor structure between the cross-cultural samples in this phase are an indication of a lack of equivalence. Importantly, this would suggest that any other comparisons between groups among the variables would be uninterpretable because obtained values would represent nonequivalent constructs. The third phase specifically assesses scaling equivalence. Scaling equivalence is determined by comparing the samples on the “true-score” units associated with each observed item of a scale (Riordan & Vandenberg, 1994). For a more thorough review of how to utilize such procedures for assessing conceptual and scaling equivalence, see Riordan and Vandenberg (1994) and Vandenberg and Lance (2000). In our review, 17% of the studies used covariance structure analysis to assess equivalence (e.g., Judge, Locke, Durham, & Kluger, 1998; Vandenberghe, Bentein, & Delhaise, 2001; Wasti, Bergman, Stenglembach, & Drasgow, 2000).

When conducting the analyses described above, researchers may find it useful to consider the threshold at which any observed differences in their underlying constructs would cause them to halt their research efforts. Initially, it may appear that our recommendations in this section fail to consider possible cases of partial invariance. Importantly, such cases may actually represent opportunities for researchers to delve into emic aspects of the different cultures in their study. Researchers may be ill advised to automatically discard measures that do not display complete invariance.

For example, suppose a study uses a 10-item instrument that is administered to samples from both the United States and Korea. A test of invariance might reveal that only 5 items are invariant. At this point, researchers might work at developing 5 emic items for each culture and adding those items to the 5 invariant items to come up with a 10-item scale in each culture (5 items being the same across the two surveys and 5 being emically distinct). This procedure would be consistent with the derived etic approach discussed earlier and is somewhat similar to Farh et al.’s (1997) approach. In summary, partial invariance should not automatically provide researchers with a rea-
son for rejecting their survey instruments. Our overall recommendations regarding this practice are for researchers to (a) conduct tests of invariance, (b) report the findings in their studies, and (c) if necessary, explain how any lack of invariance influenced them to alter items in their survey instruments.

Another statistical approach for dealing with equivalence, and for identifying items that do not function similarly across different cultures, is item response theory (Ellis et al., 1993). Item response theory is a theory-grounded process that models the distribution of respondents’ scores at the item level (Fan, 1998). This process produces item statistics independent of respondent statistics and person statistics independent of the survey items administered. This invariance property of the theory has made it possible to solve important measurement problems that have been difficult to address with other frameworks, and it has established the basis for theoretically justifying the use of item response theory models (Fan, 1998).

The models generated from this process describe the relationship between a respondent’s observable response to an item and the respondent’s standing on the unobservable trait measured by the survey instrument (Ellis et al., 1993). An item characteristic curve can then be used to display this relationship, showing the response probability as a function of the trait measured by the instrument. When item characteristic curves estimated separately for the same item for two samples are the same, the item is said to function equivalently for both groups, and when the item characteristic curves differ by more than sampling error, then there exists what is called differential item functioning (Ellis et al., 1993; Hambleton & Swaminathan, 1985; Hulin, Drasgow, & Parsons, 1983; Lord, 1980; Thissen, Steinberg, & Wainer, 1988, 1989). Differential item functioning is an indication of a lack of measurement equivalence for a particular item in a survey. Differential item functioning items should therefore not be used to compare samples in cross-cultural research because such comparisons would be based on response tendencies rather than on true differences in the construct of interest.

Only three studies in our sample (Robert, Pobst, Martocchio, Drasgow, & Lawler, 2000; Ryan et al., 2000; Schmit, Kihm, & Robie, 2000) used item response theory to test for equivalence. All three, however, provide further explanation on how this statistical technique can be used in cross-cultural research.

**Conclusion**

Increases in international business have recast a spotlight on cross-cultural research. Survey instruments commonly used in domestic contexts are now being applied in other cultural settings, and new surveys are being developed specifically for nondomestic applications. These practices have presented researchers with unique issues and problems (Saeed & Athanassiou, 1998). We mentioned in the beginning of this article that such issues are applicable to a variety of field-research settings, including single-sample domestic studies. Nevertheless, cross-cultural research may be particularly susceptible to the pitfalls we have identified. For example, the use of convenience samples can be problematic in all research contexts but may be more prevalent in cross-cultural research, especially when samples are drawn from different countries. Under such conditions, identifying key characteristics to match samples on (or to use as control variables) may be relatively more challenging.
Similarly, problems related to procedural equivalence (in terms of survey administration) are important for all studies using multiple samples but may be particularly important in cross-cultural applications due to the differences across cultural contexts that were discussed earlier. These include language differences, variability in the perceptions of researcher intrusiveness, and respondents’ levels of familiarity with filling out surveys. Such differences contribute to situations that would make cross-cultural researchers more prone to procedural inconsistencies.

Issues related to semantic, conceptual, and scaling equivalence, although important in all types of research, may also be particularly pressing in cross-cultural research. Respondents across different cultural contexts are more likely to carry different internal meanings and frames of references associated with key constructs and may also be subject to different norms concerning scale ranges and acceptable response patterns on survey instruments.

Finally, the Western bias that seems to be prevalent across many U.S. contexts also highlights the need for cross-cultural researchers to be especially attentive to the issues we have presented (Dueck, 1983; Hofstede, 1991; Tapp, 1981). Because of such bias, there is more of a chance that survey administration procedures, levels of rapport, instructions, translations, item phrasings, and theory in general may be rooted in U.S. culture (there may also be similar biases that stem from host-nation cultures when non-U.S. researchers conduct cross-cultural research in foreign settings). We addressed methods for handling such problems throughout the article, including the use of derived etics and the use of both insiders’ and outsiders’ perspectives. Such practices are related to what others have called “decentering,” a process in which researchers from different cultures develop research questions and instruments out of different cultural environments (see Berry, 1980; Hofstede, 1991).

In sum, the issues presented in this article, although applicable to all researchers, represent particularly relevant concerns for cross-cultural researchers and should be addressed in the design and implementation phases of their studies.

In this review, we have presented three stages of the research process, and for each stage, we have identified some best practices that are meant to deal with cross-cultural complexities. These recommendations are summarized in Table 3.

We hope that as researchers continue to explore the suitability of using their theories and constructs across cultures, they will use these best practices as a checklist for verifying the validity and methodological soundness of such applications. That being said, we recognize that the unique challenges faced by cross-cultural researchers will sometimes prevent them from adhering to each recommendation we have put forth. For example, we stated that country should not be used as a proxy for culture, but we also recognize that there may be legitimate constraints on data collection that limit opportunities to use other delimiters of culture. This type of limitation should not automatically discourage researchers from pursuing their initial goals, especially if their studies have other valuable features (e.g., data being collected from 20 or 30 nations). In such cases, we encourage researchers to recognize the methodological issues faced in their particular studies, discuss the ways they attempted to address them, and if they were unable to address them or overcome them, recognize them as part of the studies’ limitations. Such limitations represent valuable opportunities for future endeavors and in this way can contribute to the growth and development of cross-cultural research.
Appendix

Table 3
Summary of Best Practices Identified

<table>
<thead>
<tr>
<th>Development of the cross-cultural research question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use an emic-etic (derived etic) approach</td>
</tr>
<tr>
<td>Incorporate culture into the theoretical framework</td>
</tr>
<tr>
<td>Use other delimiters besides country to operationalize culture</td>
</tr>
<tr>
<td>Measure Hofstede's cultural value dimensions directly in the specific research context</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alignment of the research contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equivalence of samples</td>
</tr>
<tr>
<td>Match samples</td>
</tr>
<tr>
<td>Statistically control for sample differences</td>
</tr>
<tr>
<td>Explicitly describe sample characteristics in studies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Administration of surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implement uniform survey administration procedures</td>
</tr>
<tr>
<td>Provide explicit instructions and examples in a consistent fashion across samples</td>
</tr>
<tr>
<td>Establish levels of rapport while maintaining standardization and consistency</td>
</tr>
<tr>
<td>Explicitly describe the procedures used to establish uniformity in survey administrations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validation of the research instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic equivalence</td>
</tr>
<tr>
<td>Use back-translation</td>
</tr>
<tr>
<td>Avoid the use of figures of speech, terminologies, and phrases that are common to only one culture</td>
</tr>
<tr>
<td>Be cognizant of words/phrases that elicit cognitive or affective states</td>
</tr>
<tr>
<td>Use surveys in pilot studies</td>
</tr>
<tr>
<td>Consider both insiders' and outsiders' perspectives</td>
</tr>
<tr>
<td>Explicitly describe the procedures used to establish semantic equivalence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scaling and conceptual equivalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use covariance structure analysis</td>
</tr>
<tr>
<td>Use item response theory</td>
</tr>
</tbody>
</table>

Appendix

Articles Examined for Best Practices


### Note

1. The studies reviewed in this article have authorship from a variety of countries. In addition, the journals from which these studies were drawn each have editorial boards with representation from three or more countries. Therefore, while many of the journals are American based, the cross-cultural research within them is thought to be representative of both American and non-American perspectives.

### References


Bryan S. Schaffer is a doctoral candidate in organizational behavior and human resource management at the Terry College of Business, University of Georgia. He received an M.B.A. from Georgia State University in 1989. His research interests include cross-cultural studies, organizational diversity and relational demography, and methodological issues in cross-level research. He has accepted a position as an assistant professor of management at the University of North Carolina–Asheville, beginning August of 2003.

Christine M. Riordan is an associate professor in the Department of Management at the Terry College of Business, University of Georgia. She is also the director of the Institute for Leadership Advancement in the Terry College of Business. Her current research includes the study of the effects of human resource practices on group and organizational performance and labor force and cross-cultural diversity. She has presented numerous papers on these topics at national and regional conferences and has published her research in a variety of journals.