Half a century of measuring culture: Review of approaches, challenges, and limitations based on the analysis of 121 instruments for quantifying culture

Vas Taras [a,*], Julie Rowney [b,1], Piers Steel [c,2]

[a] University of North Carolina at Greensboro, Bryan School of Business and Economics, 349 Bryan, POB 26165, Greensboro, NC 27402-6165
[b] University of Calgary, Haskayne School of Business, Human Resources and Organizational Development, 448 Scurfield Hall, 2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4
[c] University of Calgary, Haskayne School of Business, Human Resources and Organizational Development, 444 Scurfield Hall, 2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4

ARTICLE INFO

ABSTRACT

After examining 121 instruments for measuring culture, we provide a historical overview and analyze how culture has been operationalized over the last half a century. Our study focuses on the topics of culture definition, dimensionality of culture models, collection and analysis of data for measuring culture, levels of culture measurement, issues of cross-cultural survey equivalence and the reliability and validity of culture measures. For each of these topics, we provide a review of existing approaches, discuss the challenges, and suggest best practices. Based on our analysis, we identify gaps in the field of culture measurement and offer directions for future research.

© 2009 Elsevier Inc. All rights reserved.

"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be."
Lord Kelvin, 1883

Culture is an extremely popular construct. A Google search for "culture" provides over half a billion hits and the number of hits using Yahoo! is over two billion which is more than for "politics", "war", "environment", or "sex."[3] Major social science electronic databases provide links to 100,000 to 700,000 articles when "culture" is used as the search key word. Although the concept of culture is centuries old, first attempts to quantify it were not undertaken until the middle of the twentieth century. Since then, culture measurement techniques have improved dramatically as more than a hundred instruments have been developed and hundreds of studies involving quantitative assessment of different facets of culture have been published.

Although attempts to quantify various aspects of culture can be traced further back in time (e.g. England, 1967; Haire et al., 1966; Kluchhohn and Strodtbeck 1961; Kuhn and McPartland 1954; Rokeach 1973), it was not until the publication of Hofstede's "Culture's Consequences" in 1980 that we experienced an explosion of interest in the issue of culture measurement. The popularity of Hofstede's IBM study can be partly credited to its large international sample and to the fact that it was the first one to employ

---

* Corresponding author. Tel.: +1 336 256 8611.
  E-mail addresses: v_taras@uncg.edu (V. Taras), julie.rowney@haskayne.ucalgary.ca (J. Rowney), Piers.Steel@Haskayne.UCalgary.ca (P. Steel).
[1] Tel.: +1 403 220 6592; fax: +1 403 282 0095.
[2] Tel.: +1 403 220 8428; fax: +1 403 282 0095.
[3] As of Aug 24, 2008.

relatively advanced, for its time, research designs and statistical analysis tools. The result of the study was a coherent theoretical model and a concise set of quantitative indices describing and ranking cultural values along several dimensions. Given this impact, it is surprising that the IBM study was not originally intended for the purpose of cross-cultural comparison. It was initially based on a questionnaire with almost two hundred items that dealt with attitudes to and perceptions of various organizational issues such as work satisfaction, commitment, management-related issues and others (i.e., an Employee Attitude Survey). It was during the analysis stage that several "cultural" items were selected to become what is now known as the Culture Survey Module (Hofstede, 1980, 2001).

Following the IBM study, several alternative models of culture and their corresponding instruments gained recognition and popularity, such as those offered by Trompenaars (1993), Schwartz (1994), Maznevski and DiStefano (1995), Inglehart (1997) and the GLOBE team (House et al., 2004). Nevertheless, due to various reasons including restrictions on the use of the existing questionnaires and dissatisfaction with the quality of the survey tools, a great number of scholars chose to create their own alternative measures of culture, with a substantial number of them being developed in recent years. This trend suggests that in the years to come more models of culture and tools for measuring culture will be offered.

Through a comprehensive literature search, we identified 121 instruments that have been used in the quantitative measurement of culture.[4] This paper is based on a detailed analysis of the methods, definitions, and original items utilized in these instruments. Of note, though the majority of the instruments included in the review were specifically designed to measure culture, a few of them were originally introduced as measures of various social or psychological constructs (e.g., Carver et al., 1989; Kuhn and McPartland, 1954; Spector, 1988), but have subsequently been used for cross-cultural comparisons. This suggests that these constructs may also represent dimensions of culture.

Building upon our analysis of the 121 instruments, this study provides a review of approaches to quantifying culture, discusses challenges, and suggests best practices for measuring culture. The list of discussed topics covers the complete range of issues related to culture measurement, including the definition of culture, dimensionality of culture models, survey item composition, data collection for quantifying and comparing cultures, measurement equivalence, levels of measurement, and the reliability and validity of culture survey instruments. Based on the review of the literature and the existing culture measurement instruments, we identify trends in the area of culture measurement and analyze the factors that shape the development of the field. Many of our arguments rely on results of a survey of expert opinions conducted specifically for this study. Based on our analysis, we identify gaps and areas that have not been sufficiently explored and suggest a list of fundamental questions that need to be addressed by future research involving culture measurement.

1. Defining culture

Like most common language words drafted into scientific study, there is no commonly accepted definition of the word "culture." More than 50 years ago, Kroeber and Kluckhohn (1952) found 164 distinct definitions of culture. More recent reviews indicate that the number of definitions are increasing (Hofstede, 2001; Soudijn et al., 1990). Moreover, it appears that the construct of culture has been studied in many disciplines under different names. Koltko-Rivera (2004) offered an extensive review of the research on worldviews and pointed out the existence of parallel redundant research streams that essentially explore the same phenomenon but use different terminology. Based on the review of Koltko-Rivera (2004), there are no substantive differences between constructs that have been researched in different subfields of anthropology, psychology, sociology, and management as "culture," "value orientations," "basic beliefs," "schemas," "philosophy of life," "world outlook," and many other names, and thus the number of definitions of the construct may be much greater than any single discipline can offer.

Despite the great complexity of the construct of culture, or possibly because of it, scholars often define and operationalize culture overly simplistically, which can lead to problems. There are numerous examples when culture is simply equated to nationality or citizenship, as evident from such quotes as "cultural background was measured by the current citizenship (passport status) of each of the managers" (Offermann and Hellmann, 1997: 346), "Individualism–collectivism was operationalized by the respondent's native culture (country of origin)" (Trubisky et al., 1991: 73), or "participants were divided into high and low Power Distance groups by country-of-origin" (Eylon and Au, 1999: 378). According to Schaffer and Riordan (2003), approximately 79% of cross-cultural studies published between 1995 and 2001 used country as a proxy for culture.

Based on the analysis of the available definitions of culture, we have found that while the existing definitions vary greatly, there are several common elements present in virtually all of them. First, it is generally agreed that culture is a complex multi-level construct. It is often depicted using an "onion" diagram with basic assumptions and values representing the core of culture, and practices, symbols, and artifacts representing the outer layers of the construct (e.g., for different versions of the "onion" diagram see Hofstede, 1980; Trompenaars, 1993). Second, culture is shared among individuals belonging to a group or society. Third, culture is formed over a relatively long period. Finally, culture is relatively stable. This is where the agreement ends. Different models of culture and consequently different instruments for measuring culture focus on different levels (e.g., national, organizational, individual) and aspects (e.g., values, practices, observable artifacts and rituals, underlying implicit assumptions).

Culture scholars from different fields tend to focus on different elements of culture. Traditionally, culture has been in the domain of anthropology and archaeology where it has been defined and studied, mainly qualitatively, with the emphasis on the

---

[4] Refer to the Culture Survey Catalogue (Taras, 2008) for the complete list of reviewed instruments.

external layers of culture such as artifacts, languages, and traditions. With an increase in immigration and the cross-border expansion of business, cross-cultural issues have also become salient to management, psychology, and education. Following the anthropological tradition, most of the early social science studies of culture were qualitative and also defined and studied culture by focusing on artifacts and traditions. Their scope was largely limited to descriptions of protocols, customs and the ways of doing business in certain societies.

With the introduction of the early quantitative models of culture, the focus shifted from artifacts and cultural practices to cultural values and attitudes that presumably govern human behavior. Although directly observable cultural artifacts have always been recognized as an important aspect of culture, being described predominantly qualitatively limits their use in contemporary scholarly journals that tends to favor quantitative approaches. Consequently, values have become the focal point in cross-cultural literature. As a result, the terms "cultural values" and "culture" have often been used interchangeably, especially in the fields of management and psychology where the influence of the "Culture's Consequences" of Hofstede (1980) has been most obvious.

Nevertheless, there is no single opinion as to what values and attitudes constitute culture. Despite the overlap in dimensions from different models of culture, there are differences in the types of values and attitudes that are emphasized in different fields. For example, management scholars are mainly concerned with work-related values (e.g., Hofstede, 1980), their colleagues in sociology and social psychology compare cultures by attitudes to social and political issues (e.g., Inglehart et al., 2004), and psychologists call attention to cross-cultural differences in self-perception (e.g., Singelis, 1994).

Although values have been the focal point of cross-cultural studies in recent decades, a number of models attempted to go beyond the value-based paradigm of culture. Some early attempts had been undertaken by Hall (1977) who classified culture into low and high-context groups based on prevailing communication style and pattern. Unfortunately, the model has remained only a theory. It is believed that Asian cultures tend to be high-context oriented, while the opposite is true for Western societies, but no large-scale empirical cross-cultural comparison study has been undertaken to quantitatively position existing cultures along the low–high-context dimensions. Another value-free model of culture was offered by Gelfand et al. (2006). The model is based on the notion of cultural looseness–tightness; that is the classification is made in terms of the degree to which social norms and sanctions are enforced within societies. Although no quantitative national indices based on the model have been published yet, the data collection efforts are currently in progress and the conceptual model may soon be supplemented with empirical findings. Finally, Bond, Leung and their colleagues (Leung and Bond, 2004; Leung et al., 2002) offered a model that goes a level deeper than values and describes cultures in terms of their "social axioms" or basic assumptions about social complexity, spirituality, perceived fate control, cynicism, and rewards. The model has been also supported by empirical data from over forty societies (Bond et al., 2004).

To sum up, it has been recognized that culture is a multi-level, multi-facet construct. It has been generally agreed that culture is distinctively different from personality or individual temporal states as it is a group phenomena. For the purpose of the current study, our working definition of culture is based on these common elements: culture is a group's shared set of distinct basic assumptions, values, practices, and artifacts that are formed and retained over a long period of time.

1.1. Defining culture: challenges and best practices

A precise measurement of a construct is impossible without having a clear idea about what exactly it constitutes. Unfortunately, defining the boundaries of culture and separating it from related constructs has proven to be a challenging task. For example, a number of scholars have called for a closer look at the substantive similarities between culture and personality (Benet-Martinez, in press; Benet-Martinez and Karakitapoglu-Aygun, 2003; Hofstede and McCrae, 2004; Steel and Ones, 2002). There may be a very fine line between cultural and personality or, in other words, between values determined by cultural background and values that possibly have less to do with culture and are strictly personality traits. Consider "anxiety." Is it a measure of the cultural construct of Uncertainty Avoidance (Hofstede, 1980) or does it represent the personality facet of Neuroticism (Costa and McCrae, 1992)?

Multiple empirical studies have found rather high correlations between personality traits and various value systems (e.g., Furnham, 1984; Hofstede and McCrae, 2004), which raises the question of how distinct these two domains are. Some recent studies looked at the geography of personality traits (Allik and McCrae, 2004) and found cross-national differences in personality profiles. If a personality trait prevails in some ethnic group, does it make it a facet of culture? Where does culture end and other constructs, for example personality, begin? It appears that even though there are obvious conceptual differences between personality and culture, a comparison of culture and personality measurement instruments reveal very little methodological and empirical difference.

With the focus on cultural values, published research commonly neglected the existence of other important layers of culture. It has generally been assumed that those are the values that govern behavior and thus the other components of culture are of lesser importance. However, the strength of the relationship between different layers of culture is still unclear and thus neglecting other aspects of culture, such as basic assumptions, communication styles, cultural looseness–tightness, as well as artifactual elements of culture, limits applicability of results of cross-cultural studies. There have been a few efforts to explore culture beyond values (e.g., Gelfand et al., 1998; Hall, 1977; Leung and Bond, 2004) but more theoretical and empirical research is needed in this direction.

A single model cannot comprise all aspects of such a highly complex, multidimensional and multi-layered phenomenon as culture. A single numeric index or a few dimension scores cannot provide a comprehensive description of culture. The nature of the relationship between different elements of culture is still to be determined and one must be very cautious about drawing parallels and generalizing findings across culture facets (e.g., language, values, practices) and levels (e.g., individual, national). Thus, it is very important to specifically define which elements of culture are the focus of a model and avoid unjustified generalizations of the findings to facets of culture that are not directly measured in the study.

2. Dimensions of culture

Based on our review, all existing known models of culture and their corresponding survey instruments are multidimensional. A typical model of culture is usually represented by a set of dimensions that capture a range of cultural values, attitudes or practices. Usually an instrument measures culture by quantifying values, assumptions, or practices along four to eight cultural dimensions or factors. We conducted a comprehensive literature review for existing instruments for measuring culture and their underlying theoretical models. The search included all available sources, including scholarly journals, books, electronic databases, theses, conference presentations, and personal archives, and was not restrained to particular journals or fields of study. Our literature review identified 121 distinct instruments for measuring culture (for a complete list and details on each instrument see Taras, 2008).

Table 1 provides a summary of specific types of measures that were included in these reviewed instruments. The measures were grouped based on a comparison of names and definitions of specific dimensions included in each of the reviewed instruments, as well as an analysis of the original items included in the reviewed instruments. This analysis identified the twenty-six most popular facets of culture (Table 1).

The influence of the "Culture's Consequences" of Hofstede (1980) on the field is obvious. His framework has been favoured by cross-cultural scholars for a number of reasons, including limited availability of alternatives, convenience, popularity, and simply habit. The dimension names originally introduced by Hofstede have been repeatedly used in subsequent studies. Furthermore, the twenty-six facets we identified can be grouped, with a few exceptions, into four major blocks related to Hofstede's dimensions of Individualism–Collectivism, Masculinity–Femininity, Uncertainty Avoidance, and Time Orientation. The grouping has been done based on a construct, content, and criterion analysis. A thorough comparison of dimension definitions and items in the reviewed instruments and those offered by Hofstede, as well as very high correlations between Hofstede's scores and those obtained using other instruments (for review see Hofstede, 2001, pp. 220–223) confirmed their close conceptual and empirical correspondence.

Based on our review, 51.2% of the models contain unique dimensions, such as Universal–Particular or Affective–Neutral in the model of Trompenaars (1993), Hedonism and Benevolence in the model of Schwartz (1994), and Determinism and Fate in the model offered by Maznevski and DiStefano (1995). However, 97.5% of all reviewed measures contain at least some dimensions that are conceptually similar to those introduced by Hofstede. Very few models, normally those from non-management literature, were found to contain no links to Hofstede's work. For example, the dimensional structure of the World Value Survey (Inglehart et al., 2004) is completely different from Hofstede's model, though empirical relationships between the models are quite strong (Hofstede, 2001).

Table 1
Measures of culture included in the reviewed instruments.

Measure	Dimension	Instruments containing the measure/dimension			
		Number [a]		%	
Self vs. group interest	Individualism–collectivism	36	79	29.8	65.3
Group loyalty		36		29.8	
Teamwork and cooperation		35		28.9	
Self-reliance		35		28.9	
Family integration		28		23.1	
Conformity		24		19.8	
Equality seeking		21		17.4	
Self-perception		19		15.7	
Social responsibility		17		14.0	
Personal independence		16		13.2	
Self-identity		10		8.3	
Achievement/competitiveness	Masculinity–femininity	28	36	23.1	29.8
Assertiveness		10		8.3	
Confrontation avoidance		8		6.6	
Gender equality		4		3.3	
Accepted inequality	Power distance	20	23	16.5	19.0
Acceptance of/preference for authoritative decision making		18		14.9	
Ambiguity avoidance	Uncertainty avoidance	9	11	7.4	9.1
Risk avoidance		2		1.7	
Long vs. short-term orientation	Long- vs. short-term orientation	5	7	4.1	5.8
Tradition		5		4.1	
Emotional vs. neutral	Unique measures	6	62	5.0	51.2
Pleasure-seeking		3		2.5	
Relationship to environment		3		2.5	
Determinism/fate		2		1.7	
Machiavellism		2		1.7	
Other		59		48.8	

[a] Indicates the number of instruments in which the measure is present.

Several models further refine their factor structures by using subdimensions. For example, the Individualism–Collectivism measures offered by Hui and colleagues (Hui, 1988; Hui and Triandis, 1986; Hui and Yee, 1994) contain items aligned along several subdimensions such as Spouse, Mother, Sibling, Relative, Friend, Co-Worker, Neighbor, Acquaintance, Stranger, and Foreigner. In most models, cultural constructs are unidimensional and bi-polar. For example, Hofstede measures attitudes to ambiguity by a single bi-polar dimension of Uncertainty Avoidance, with high and low Uncertainty Avoidance as the extremes. Similarly, the construct describing relationships to a group is represented by a single dimension with Individualism and Collectivism as its poles. In contrast, a number of instruments contain multidimensional cultural constructs. For example, in some models, the construct of Individualism is represented by multiple dimensions such as Vertical and Horizontal Individualism and Collectivism (Singelis et al., 1995), or Allocentrism and Idiocentrism (Triandis, 1994b). Similarly, in the model offered by Maznevski and DiStefano (1995), Relationship to Nature is measured along three dimensions: Mastery, Subjugation, and Harmony.

Various instruments use different methods for combining individual item scores into dimension totals. Some older instruments, such as Hofstede's Value Survey Module (VSM), employed complex equations with weightings and constants to obtain final dimension scores. Because of the complexity and a lack of theoretical justification for varying item weightings and constants, the approach never gained popularity. Instead, the dimension scores are usually calculated as the average or the sum of the item scores comprising the dimension. For instruments employing scenario items, the test score is usually represented by the percent of people choosing certain nominal categories (e.g., Triandis et al., 1998). Similarly, for sentence completion instruments, the final score is usually the percentage of responses classified as, for example, individualistic (Kuhn and McPartland, 1954). Finally, for value ranking tests, the final score is usually represented by the position number or a corresponding value (Rokeach, 1973). However, this type of survey tool is usually used when the focus is on the value structure and not on the final score.

Once the dimensions are chosen, the next task is to define the boundaries of each dimension and decide on the list of specific measures representing each dimension. The review indicates that items referring to different elements of culture have often been mixed within a single set leading to possible construct contamination. Here is an example from our sample of survey instruments (Chew, 1996): "I avoid engaging in recreational activities that could potentially result in injuries" (practice) and "To risk one's life for adventure is to affirm one's existence" (belief). While it has been traditionally assumed that there is a positive relationship between different layers of culture (e.g., values and behaviors) it is not necessarily true. In fact, a possibility of no relationship or even a strong negative relationship between values and practices has been empirically demonstrated by the GLOBE study (House et al., 2004).

The reason for the frequent questionable integrations of seemingly different measures into a single construct can be partly explained by the confusing construct definitions in the "Culture's Consequences" of Hofstede (2001) and the interpretations and generalizations that Hofstede made based on his findings. In the case of Individualism–Collectivism, the different versions of Hofstede's survey tool (VSM) operationalized the construct through a set of items that referred to the desired area of residence, concern for working conditions, employment security, need for the adventure element in the job and cooperation at work. At the same time, the construct definition in the first edition of "Culture's Consequences" (Hofstede, 1980) rather vaguely stated that the dimension "describes the relationship between the individual and the collectivity [and] the way people live together" (p. 148). Even broader was Hofstede's extended list of suggested implications of cultural differences along the Individualism dimension. For example, he elaborated that in individualistic societies, initiative is encouraged, job autonomy and challenge are aspired, and people tend to pursue their own ends without minding others. Other generalizations dealt with conformity, family-related issues, involvement in organizational life, and teamwork. Because of the iconic status of Culture's Consequences and the frequent misinterpretations of Hofstede's findings, it has often been taken for granted that each of these elements is a facet of the single construct that Hofstede called Individualism–Collectivism.

Another construct that often causes confusion is that of Masculinity–Femininity. Many instruments for measuring the construct include items referring to such theoretically separate factors as Achievement Orientation and Gender Egalitarianism. For example, the measure of Masculinity developed by Furrer et al. (2000) contains items referring to both factors ("Money and material things are important" and "Both men and women are allowed to be tender and to be concerned with relationships"). The confusion may have been caused by the varying definitions of the construct provided by Hofstede (compare 1980, 1984, and 2001) and the wide range of characteristics he posited about the cultures that represent different poles of the dimension. For example, in his definitions of Masculinity, Hofstede focused on achievement orientation. However, further in the text he supplemented his definition with a set of assumptions about the difference in gender roles and the extent to which gender roles are defined in masculine versus feminine societies. Other examples of commonly confused constructs are Uncertainty Avoidance with Risk Aversion and Power Distance with Power Seeking.

A separate note must be made on the emic vs. etic approach to defining and measuring culture. The terms' origin is linguistic. "Etic" refers to sounds common to all languages and "emic" indicates sounds specific to one language or a group of related languages (Gannon, 2004). In cross-cultural studies, the etic approach assumes that there is a set of universal cultural dimensions that are equally relevant to all cultures. A good example of this approach is the model of Hofstede (1980) that is based on five cultural dimensions that are used to classify and cluster a wide range of national cultures. According to Hofstede (1980, 2001), the dimensions are universal and any society can be positioned relative to other societies along each of the dimension continuums. On the contrary, the emic approach is based on the assumption that at least some cultural dimensions are culture-specific and cannot be used to analyze cultures of different societies. Unlike universal cultural dimensions that are used to describe cultures relative to each other, emic studies are unique to specific cultures and are used to describe each culture independently.

Virtually all reviewed models and instruments for quantifying culture use the etic approach. However, a few exceptions were found. Cultural metaphors are a good example of the emic approach. Although emic cultural metaphors were used in cross-

cultural studies for the first time more than 30 years ago (Geertz, 1973), the etic approach has dominated the field of cross-cultural studies and it was not until the 2000's that the emic approach started to gain popularity again (e.g., Denny and Sunderland, 2005; Gannon et al., 2005–2006; Paulson, 2005). Typically, the cultural metaphor approach is qualitative, but quantitative analysis is also possible under this paradigm. Most of the proponents of the emic approach, however, do not limit their analysis to cultural metaphors and believe that the emic and etic approaches must be integrated and used simultaneously to truly understand culture.

2.1. Dimensions of culture: challenges and best practices

As indicated, it is extremely difficult to create a model that effectively and efficiently captures all aspects of such a complex phenomenon as culture. The number of dimensions has been growing from four, later five, in Hofstede's original model to seven in the model offered by Trompenaars (1993), ten in the model offered by Schwartz (1994), six tri-dimensional constructs in the model offered by Maznevski and DiStefano (1995), and eighteen dimensions (nine dimensions of cultural values and nine dimensions of practices, each measured separately at the organizational and national levels) in the GLOBE model of culture (House et al., 2004). However, despite the great variety of dimensions, it is still too early to claim that every aspect of culture is captured by any single model or even by all existing models taken together. As noted by Bing (2005), it is tempting to claim that the questionnaires and associated databases provide the coordinates for the entire map of culture, but we do not yet know the complete map of culture, nor are we likely to in the near future.

On the other hand, while the wide array of measures can presumably provide a richer description of the studied phenomenon, the marginal utility of additional measures is probably diminishing. Keeping a balance between comprehensiveness and parsimony is very important. Additional measures of culture become at some point redundant and make a model unworkably bulky. Furthermore, should we attempt to include the entire range of options in a model of culture? Are all values (or sources of guidance, or practices) determined by culture? For example, are attitudes towards teamwork (Wagner and Moch, 1986) supposedly a cultural construct, determined by one's cultural background or are they independent of culture? Obviously, simply finding a statistical difference between two national averages is not sufficient to conclude that the construct is a facet of culture or is culture-bound.

Because we could find no empirical studies that provided a well-substantiated answer to the question, we decided to shed some light on the issue by gathering expert opinions. We put together a comprehensive list of cultural dimensions found in the 121 analyzed instruments for value quantification and sent it to thirty-six leading cross-cultural management scholars. Using a seven-point Likert scale, the participants were asked to evaluate, based on their experience, the extent to which each of the dimensions on the list were determined by culture, as well as the extent to which they affect individual behavior in the workplace. The survey was anonymous and the responses were collected using an on-line survey service. The culture experts were selected using the following procedure. For a different meta-analytic project, we conducted an extensive search and identified 508 empirical studies that involved the quantitative measurement of culture. The authors of the studies were ranked by the number of publications and those who authored or co-authored four or more papers (thirty-seven scholars) were identified as leaders in the field of cross-cultural studies and contacted with an invitation to complete the survey. Virtually all of them were university professors specializing in cross-cultural research in various subfields of management, psychology, and sociology, and currently working in twelve different countries. Twenty-eight scholars anonymously responded to our call (77.8% response rate). The results are presented in Table 2. Of note, the list of the dimensions in Table 2 somewhat differs from those in Table 1, because the latter includes a list of specific measures that were used to evaluate cultural values along each of the dimensions.

According to the results of the survey, quite a few dimensions on the list scored low on the extent to which they are perceived to be related to culture. Among them are Believing in Evil/Good and Changeable/Unchangeable Basic Human Nature, Inclination to Teamwork, Machiavellism, Pleasure-Seeking (a.k.a. Hedonism), and Risk Avoidance. Of course, these results are a mere reflection of subjective options; nevertheless, they may be of interest to those who intend to develop a new instrument for quantifying culture.

Another related challenge is choosing the cultural dimensions that are most relevant to the research questions of a given study. For example, a management scholar may be interested only in the cultural values that directly affect workplace behavior. For this reason, he or she may want to exclude the dimensions that are describing attitudes to, for example, political or family-related issues. Unfortunately, as in the case of culture/personality differentiation, it may be difficult to determine which values are most relevant to a given phenomenon. As per the summary in Table 2, only some values were classified by the leading cross-cultural management scholars as relevant to the workplace. Quite a few of the popular culture measures obtained rather low relevance rankings.

Based on the results of the expert opinion survey, very few dimensions from our lengthy list could be classified as both determined by culture and related to the workplace. Among them are Ambiguity Avoidance, Gender Equality, Power Distance, Status Ascription, and Universalism–Particularism. Unexpectedly, the measures related to the immensely popular Individualism–Collectivism construct did not score highly, with the possible exception of Independent–Interdependent Self-Perception that came out in the sixth position with a combined score of 7.14. Evidently, it is not always better to include more dimensions in the theoretical model and subsequent analysis. Depending on the nature of the research questions, the scholar must choose the dimensions that are most relevant for the study.

Once it has been agreed as to how culture is defined in a specific research project and what dimensions of culture should be measured, the next hurdle is defining the boundaries of specific cultural dimensions. As noted by Cook and Campbell (1979), valid

Table 2
The Results of Expert Opinion Survey.

Dimensions	Relevance to workplace	Relevance to culture	Combined score
Achievement orientation: Willingness to win advance get a better position earn more and succeed at any cost even if it means harmed interpersonal relations and/or sacrifice of personal life	3.14	3.87	7.01
Ambiguity avoidance: The degree to which people are made nervous by uncertain situations and ambiguity and prefer to have clear rules guidance bureaucratic practices and rituals for every situation	3.61	3.73	7.33
Assertiveness: The extent to which an individual exhibits assertive self-confident and tough behavior and values	3.87	3.25	7.12
Attitude to ritual suicide: The degree to which individuals can accept or even encourage suicide performed as a point of honor or for a perceived higher purpose (e.g. suicide bombing or hara-kiri)	1.85	3.76	5.62
Believing in evil/good basic human nature: Believing that people are essentially bad or good and as a result always expect people to behave badly (avoid work steal lie) or well (work hard be helpful be honest)	3.09	2.91	5.99
Believing in changeable/unchangeable basic human nature: Believing that people's nature/character does not change with time	3.11	2.89	6.00
Conformity: The degree to which individuals restrain their actions inclinations and impulses that are likely to upset or harm others	3.32	3.45	6.77
Conservatism: The degree to which people resist quick change and try to preserve the traditional way of doing things	3.29	3.49	6.78
Determinism: The degree to which people believe that their paths are predetermined by the forces they cannot control and what has to happen will happen regardless of their efforts	2.64	3.45	6.10
Family integration: The degree to which individuals maintain close ties with their extended families consult their family members when making important decisions and believe that family members should live as close to each other as possible	2.70	4.17	6.87
Gender equality: Perceiving roles and abilities of men and women as equal and believing that men and women have the same rights and responsibilities and are capable of performing equally well on most work-related tasks including managing people	3.61	4.17	7.78
Pleasure-seeking: The extent to which people emphasize pleasure and enjoyment of life and attribute secondary role to the work life; belief that people work to live not live to work	3.09	2.88	5.97
Humane orientation: The degree to which individuals encourage and reward fairness altruism generosity caring and being kind to others	3.15	3.14	6.29
Inclination to teamwork: Preference to work in a team rather than work alone willingness to share responsibilities rewards and punishments with the team members and being ready to put interests of the team before personal interests	3.87	3.08	6.95
Independent–interdependent self-perception: The extent to which individuals include close relationships in their representation of self. In other words, the extent to which individuals feel that their relatives, friends, and organizations they belong to are an important part of themselves	3.46	3.68	7.14
Emotional vs. neutral: The degree to which people believe that displaying feelings at work is unprofessional and inappropriate	3.25	3.62	6.87
Machiavellism: The extent to which a person is manipulative deceiving and willing to use dirty tricks when dealing with others	3.62	2.10	5.73
Personal independence: The degree to which individuals value their privacy believe what happens to them is their own doing and prefer to struggle through personal problems or enjoy personal achievement by themselves	3.36	3.51	6.86
Power distance: The extent to which people expect and accept that power in organizations is distributed unequally; degree of inequality among people which the individual accepts as normal	3.91	3.99	7.90
Relationship depth: The degree to which individuals develop close relationships with their co-workers and remain close friends and interact frequently outside the workplace settings	3.41	3.26	6.67
Relationship to environment: Subjugation vs. mastery; the extent to which people feel they can and should change the environment vs. they should adjust themselves to the environment	3.28	3.09	6.37
Risk avoidance: The extent to which people are reluctant to take risk or make risky decisions	3.80	2.70	6.50
Self-identity: The extent to which individuals emphasize their personal identity independent of others strive to be original and different and do not like to be identified with their groups such as families or organizations	2.96	3.28	6.25
Self-reliance: The degree to which individuals tend to rely on themselves in difficult situations rather than expect help from others	3.65	3.35	7.00
Short- vs. long-term orientation: Seeking quick gains even if it means losses in the future vs. focusing on the future outcomes and being ready to suffer losses in the short-run for the sake of future gains	3.57	3.13	6.70
Status by ascription vs. status by achievement: Perceiving status based on who the person is (son of a famous or wealthy person royalty older person man) vs. perceiving status based on person's personal achievement and skills	3.61	4.07	7.68
Universalism vs. particularism: People with universalism orientation believe that rules must apply equally to everyone and under every circumstance; while people with particularism orientation believe that some exceptions from rules can be made depending on the person (e.g., close friend) and mitigating circumstances	3.65	3.73	7.38

Results are measured on the 5-point Likert-type scale, 5=very relevant, 1=very irrelevant.

and efficient facet operationalization is extremely difficult in practice. In most cases, it is impossible to describe the facet boundary in black and white. Undoubtedly, there will be some elements in the gray area that the researcher will not be able to easily classify as those that must be included or excluded from the construct measures. It is tempting to include all seemingly relevant measures

in the assessment instrument to make certain that the construct is fully captured. However, researchers must be very selective to achieve parsimony and to avoid construct contamination.

The Individualism–Collectivism of Hofstede (1980), the most popular cultural dimension in subsequent research, is an excellent example of such a challenge. As illustrated in Table 1, most of the value survey instruments that contained the Individualism dimension used multiple measures of the construct. Based on the face validity and results of some empirical studies, it is questionable whether the combinations of the measures actually represent a single theoretical cultural dimension (e.g., Triandis et al., 1986, 1993). Such distinct constructs as Social Responsibility, Self-Reliance, Independent–Interdependent Self-Perception, Preference to Teamwork and other constructs have been traditionally labeled Individualism and frequently treated as measures of a single facet. The soundness of this approach has been disputed. As noted by Oyserman et al. (2002), Individualism–Collectivism has been defined and operationalized in "overly broad and diffuse ways" (p. 44). As a result of lumping vaguely related items into a single construct, the measure loses its precision and often becomes useless for rigorous empirical analysis (Miller, 2002). More focused and specific measures are needed to ensure validity and accuracy of the measurement.

3. Collecting data for measuring culture

The examination of the instruments included in our analysis revealed that self-report questionnaires have been the most popular and virtually the only tool for quantitative measurement of culture. In some cases, face-to-face interviews have been used to obtain responses to survey items (e.g., Inglehart et al., 2004); however, the method is not conceptually different from the pencil-and-paper self-report survey approach. Only one study was found to quantify culture based on direct observation or the analysis of secondary indicators of human behavior (Vandello and Cohen, 1999).

The sample size and sampling procedures have varied greatly across studies. A few major large-scale culture-comparison projects were generally based on data from 40 to over 60 countries/societies representing the majority of the world's population (Hofstede, 1980; House et al., 2004; Inglehart, 1997; Schwartz, 1992; Smith et al., 1996), while the rest of the reviewed studies were based on data from 2 to 10 countries. Likewise, the reviewed models of culture and their corresponding survey instruments have been developed and validated with sample sizes ranging from about 100,000 individuals[5] in the infamous IBM study of Hofstede (1980), to 99,000 in World Values Survey (Inglehart, 1997), to 17,370 in the GLOBE project (House et al., 2004), to 8841 in the study conducted by Trompenaars and his colleagues (Smith et al., 1996),[6] to 6849 in the original study of Schwartz (1992) and 10,857 in his follow-up confirmatory study (Schwartz and Boehnke, 2004), and to just a few hundred in most other culture measurement and comparison studies. Of note, although some of these total sample size figures are quite impressive, the sample size distributions across countries have been typically highly skewed. For example, while some countries in Hofstede's dataset were represented by thousands of respondents, samples drawn in other countries were very small, such as 76 for Thailand or 80 for Taiwan. Most other large-scale studies are subject to the same limitation.

Based on the descriptions of the sampling procedures in the papers that introduced and described the reviewed instruments, the vast majority of studies were conducted using convenience samples. For example, Hofstede's model has been developed based on a survey of employees representing international subsidiaries of a single organization for which he worked as a consultant at the time. Moreover, over a half of the reviewed models were developed and validated using student samples (more on the advantages and disadvantages of uni-organizational research design and the use of matched student samples in cross-cultural comparison studies later). In a few studies, however, every attempt has been made to focus on a very specific theoretically-justified target group. For example, because of the study's focus on leadership and organizations, only middle-level managers were surveyed in the GLOBE project, a group that theoretically seems most relevant to the research questions of the study (House et al., 2004). Alternatively, the Schwartz's model of general cultural values was developed based on a sample of teachers. As Schwartz (1992) explains, "teachers play an explicit role in value socialization [and] they are presumably key carriers of culture" (p. 18) and hence the choice of the target group.

3.1. Collecting data for measuring culture: challenges and best practices

There are two major sampling-related challenges when comparing cultures. First is the issue of data availability. By definition, a culture-comparison study requires data from multiple societies, countries, or regions. A large and random sample representing the target population from every existing society is preferred, but collecting such data is very costly and usually impossible with the available resources. Lack of infrastructure and limited resource availability often leaves no choice but to use a convenience sample. This imposes serious limitations on the validity and generalizability of the obtained data.

The second challenge relates to the trade-off between sample representativeness and cross-sample comparability. Whereas in most comparison studies, it is desired to deal with perfectly homogenous samples with matched demographic characteristics, the matter is much more complex when comparing cultures. On one hand, matching sample characteristics across cultural regions

---

[5] There is some confusion as to the exact sample size of Hofstede's IBM study due to the fact that there were two waves of the survey and some respondents were surveyed twice. The total number of returned questionnaires was around 116,000, but the number of individuals who participated in the survey was around 90,000 (Hofstede, 2001).

[6] A sample of about 35,000 is cited in "Riding the Waves of Culture" (Trompenaars, 1993). However, in a later description of the project, the sample size is cited as 8841 (Smith et al., 1996).

allows for minimizing possible effects of external variables, which makes cultural comparison cleaner. When replying to critical remarks of McSweeney (2002) about the uni-organizational survey design of the IBM study, Hofstede (2002a) argued that sample characteristic equivalence must be sought when conducting a comparison of national cultures, as this would allow to control for the effects of culture-free factors on the responses. Indeed, multiple studies have empirically shown a relationship between individual and country characteristics and individual and national cultural values (e.g., Hofstede, 2001; Stedham and Yamamura, 2004; Taras and Steel, 2006a). For example, a strong relationship between wealth and individualism has been found at both the individual and national levels of analysis (Hofstede, 1980; House et al., 2004). Therefore, it would not make sense to conduct a comparison of national cultures by comparing a sample of wealthy people from one country to a sample of relatively poor people from another country, as it would be difficult to determine whether the differences in cultural values were due to cultural or socio-economic differences.

On the other hand, comparing samples with similar characteristics limits generalizability of the findings. While a sample with certain characteristics may perfectly represent one society, a sample with the same characteristics drawn in another country may not be generalizable to its population at all. For example, highly educated and relatively well-paid IBM employees from the study of Hofstede (1980) could have adequately represented the general population of developed countries, but this might not have been the case for developing nations. As noted by Schwartz (1994), the divergence of the IBM sample from the general population likely differed from country to country, with the discrepancy potentially "greater, for example, in the Third World nations (e.g., El Salvador, Pakistan) than in industrialized Western nations (e.g., Switzerland, United States)" (p. 91).

The right balance of sample consistency and sample representativeness must be chosen depending on the research questions and population of interest. While a high degree of sample characteristic matching, as in uni-organizational research design, has its obvious advantages, at times it is desired only as long as the employees of the organization in different countries represent their populations fairly well. Otherwise, the findings will be of little value as they will not be generalizable to the populations of interest and the comparison of cultures across samples will be meaningless. Representativeness is especially of concern in studies that focus on cultural values of entire societies or countries. However, at other times, the focus of the study is on a narrow segment of the population and a high consistency of sample characteristics across target populations is desired. For example, the culture of elite, such as top managers, may be more predictive on many relationships due to the concentration of power (e.g., foreign policy, health care mechanisms). However, here too, it must be kept in mind that demographic characteristics of the target group in other societies may be different. For example, top managers may tend to differ substantially in terms of their age and education across countries and thus matching samples on age may not be justified.

4. Types of culture survey items

It is disputable whether a self-report questionnaire is an effective way to measure culture (Schaffer and Riordan, 2003; Schwarz, 1999), but it seems to be perceived as the best tool for the task. Because alternative assessment methods, such as observation or experiment, are much more resource-demanding, the self-report questionnaire remains the most popular method of quantifying culture.

Rankings, ratings, scenarios and sentence completion have been the major types of items in culture measurement self-report questionnaires. The item ranking (rank ordering) approach has been quite rare. It was mainly utilized in some earlier instruments. For example, the Value Survey of Rokeach (1973) requires the arrangement of a set of values in the order of their importance to the respondent. However, responding to such a survey is difficult and time consuming and the method did not gain popularity. Moreover, the ranking approach yields non-parametric data that substantially limits the choice of statistical analysis techniques.

Alternatively, rating being an easier approach has been widely used in cross-cultural research. This method requires the respondent to rate the importance of each value on a Likert scale containing five to nine interval points (e.g., Schwartz, 1994) or a respondent may be asked to rate his or her agreement with a set of statements (e.g., Hofstede, 1980). Compared to rankings, ratings allow for using longer lists of values and adding or dropping items with no effect on the rest of the items on the list. Ranking requires respondents to express sharp preferences between every pair of values. On the other hand, ratings do not force respondents to discriminate among equally important values and thus may provide data that is more representative of the respondents' values.

In some cases, participants are given a set of scenarios and are asked to identify from a number of options what their response or action would be in a given situation (e.g., Triandis et al., 1998; Trompenaars, 1993). Unlike rankings or ratings, the scenario approach yields categorical data that can be quantified only through a set of dummy variables or other indirect methods. However, in some cases, the categorical response options are assigned interval values (e.g., Hofstede's VSM-80) and are used in the formulas for computation of the dimension scores.

Finally, some instruments are based on open-ended questions. For example, a popular so-called "Twenty-Statement Test" (Kuhn and McPartland, 1954) requires the respondents to finish twenty statements that start with the words "I am…". Later, the answers are classified as individualistic (e.g., I am tall) or collectivistic (e.g., I am an IBM employee) and the percentage of each type of answer is used as a quantitative measure of cultural orientation.

4.1. Types of culture survey items: challenges and best practices

Choice of the "right" type of items for a culture measurement instrument appears to be largely a choice between validity and efficiency. Research has shown that the choice of the item type is likely to affect the psychometric properties of the instrument and

ultimately the results of the cross-cultural study. Peng et al. (1997) suggest that compared to rankings and ratings scales, the behavioral scenario approach is likely the most valid when assessing cultural values and attitudes. Unfortunately, scenarios tend to be much longer and less efficient than other types of self-report survey items. Responses to scenario items are often harder to code and analyze, especially to those with open-ended questions or multiple behavioral response options. As a result, many cross-cultural scholars prefer to use short rating items in conjunction with the Likert scale.

While Likert rating scales are probably most efficient, their limitations in cross-cultural research have been widely recognized. Multiple studies have confirmed a significant effect of cultural background on response style when using Likert scales (e.g., Harzing, 2006; Hui and Triandis, 1989; Javeline, 1999; Leung and Bond, 1989). Two types of response biases are generally discussed: extreme response bias and acquiescence bias. The first one refers to a systematic tendency to over express agreement or disagreement by choosing anchors towards the ends, positive or negative, of the Likert scale. Its opposite, the central tendency bias, is a systematic preference for moderate responses, as expressed through the inclination to choosing middle anchors on the scale. Some studies have suggested that survey response style is determined by culture, that is extreme responses are favoured in some cultures, while people from other cultures persistently tend to choose middle points on the scales (Bennett, 1977; Harzing and Maznevski, 2002; Javeline, 1999; Triandis, 1994a). Acquiescence bias refers to promptness to agreement and can also be affected by culture. Some studies have shown that respondents from some cultures are more prone to agreeing with survey questions (Marin et al., 1992; Smith, 2004), which makes a direct cross-cultural comparison less meaningful if it is done strictly from a mean-comparison perspective.

Several techniques have been commonly employed to correct for these sources of bias. Combining positively and negatively-worded items in a single instrument is a simple method for correcting for acquiescence response style (Schimmack et al., 2005; Smith, 2004), but it does not correct for extreme response bias. Within-subject standardization as suggested by Hofstede (1980) and refined by Leung and Bond (1989) and Schwartz (1994) and partial correlations as described by Schimmack et al. (2005) can partially correct for both the acquiescence and extreme response biases. However, scores obtained using these procedures are not independent of each other which limits the range of statistical analyses available. Further refinements of extreme style bias correction techniques have been offered (Cheung and Rensvold, 2000; van de Vijver and Leung, 1997), but each solution has its limitations.

Alternatively, it has been suggested that rankings rather than Likert scale ratings should be used to avoid the problem of response style bias. However, compared to ratings, rankings are more difficult to work with for respondents, especially when the list is lengthy. There are often tradeoffs between validity and efficiency and it is up to the scholar to find the balance. Furthermore, rankings yield non-parametric non-continuous data that cannot be analyzed with popular statistical analysis techniques such as ANOVA or OLS.

Event count items can offer a partial solution for response style bias. Rather than asking a respondent to choose an answer on a Likert scale, the survey question can inquire about a specific number of incidents or the percentage of time the respondent behaves a certain way. For example, responses to the item "*Managers should make most decisions without consulting subordinates*" (Dorfman and Howell, 1988) that utilizes a Likert scale with "Completely Agree" and "Completely Disagree" on the extremes could be rephrased as *"As a manager, what percentage of your decisions do you make without consulting your subordinates?"* or *"Out of your last five decisions you have made as a manager, how many have you made without consulting your subordinates?"*. Similarly, the item "*I would rather struggle through a personal problem by myself than discuss it with my friends*" (Earley, 1993) could be rephrased as "*Out of the last five instances you have had a personal problem, how many times have you discussed the problems with your friends?*" to minimize the response style and scale anchor interpretation bias. Unfortunately, continuous scales are not a perfect solution as such items typically refer to behaviors. It may be difficult to phrase an attitude or belief question in such a way that it can be answered in terms of a number or percentage of instances of certain observations. However, this approach can offer a partial solution for response style bias, especially when the focus of the study is on the behavioral layer of culture.

As a final note, it has been common in cross-cultural studies to focus on group (national) means and neglect other descriptive indices. For example, the result tables of Hofstede (1980) described national cultures only by averages and did not even include measures of data dispersion, such as standard deviations or ranges. Based on a meta-analytic review of over 500 studies that empirically assessed cultural values, Taras and Steel (2005) found that while reviewed papers reported cultural mean scores, less than half of the studies provided information about variance of cultural scores within groups, and of those, only about 2% explicitly referred to the measures of variance in their discussions. Focusing solely on group mean scores severely limits findings of cross-cultural comparison studies. We encourage future research to analyze other measures of data distribution within groups such as variance, skewness, or kurtosis, as they provide useful information about the culture of the entity.

5. Cross-level and cross-culture equivalence

To reiterate, culture is inherently a multi-level phenomenon. Usually, models of culture differentiate between individual, organizational and national levels of culture. The limited generalizability of relationships found at one level to other levels (a.k.a. ecological fallacy) has been known for decades. Thorndike (1939) warned about the fallacy of imputing correlations found for groups to individuals and vice versa almost 70 years ago. For instance, a relationship between Individualism and wealth at the national level does not necessarily mean that such a relationship will be present at the individual level.

In the context of culture measurement, the problem of faulty equivalence assumptions is not limited to generalizability across levels of measurement, but also refers to generalizations across cultures. A relationship between variables found in one culture may not be generalizable to other cultures. Other closely related equivalence issues beyond relationship findings are those of data

aggregation, equivalence of instrument properties across levels and across cultures, levels of item wording and scale interpretation equivalence across cultures. We discuss each of these one by one.

First is the issue of data aggregation. Since culture, by definition, is a shared phenomenon, cross-cultural scholars are usually interested in numbers describing cultures of groups (countries, societies, etc). Unfortunately, the traditional method of data collection, the self-report questionnaire, yields individual-level data. Although different methods could be used to describe the culture of a society using responses from individuals representing the society, deriving group averages by aggregating data from the individual to the group level has been traditionally used for this purpose. We have found no culture measurement studies that used alternative group response summary techniques such as mode or median scores.

Second is the issue of factor structure generalization across levels. One of the choices that scholars attempting to develop a model of culture are faced with is whether the factor structure should be tested with the individual or aggregated data. A factor structure that has been derived using individual-level data may not hold if the researcher first aggregates individual responses to the national level and then conducts a factor analysis using data points representing countries and vice versa (Peterson and Castro, 2006). For example, Hofstede (1980) first aggregated individual responses to each question in his survey to the national level and only then explored correlations between items and derived his four cultural factors. However, the four-factor structure does not replicate with individual-level responses (Hofstede, 2002b; Spector et al., 2001). As a result, Hofstede has always emphasized that his model applies only to the national level of analysis and may not be appropriate for studying individual cultural orientations.

A related issue is generalization of survey reliability and validity across levels. For example, reliability coefficients calculated using individual-level data do not necessarily hold with the aggregated level data and vice versa. Similarly, the criterion or predictive validity of an instrument assessed by analyzing correlations of the survey scores with theoretically relevant variables may differ depending on the data level.

While Hofstede's approach (aggregation first, factor and reliability analysis next) has been favoured in some subsequent large-scale culture-comparison projects (e.g., House et al., 2004), most chose to work with individual-level data. Testing factor structures with aggregate-level data requires a large number of samples/countries; for a factor analysis, it is generally ten data points per factor, though some argue that the ratio should be thirty to one (Kim and Mueller, 1978). Thus, even for a simple four-factor model, the dataset must be represented by forty countries. Of note, factor structures of some datasets used in the development of value models were subsequently tested with aggregate-level data. Typically the findings revealed differences in dimension structures between the individual and the national level. For example, the ten-factor model originally offered by Schwartz (1992) did not perfectly replicate with aggregate data and an alternative seven-factor solution was suggested for a national culture framework (Schwartz, 1994). The individual-level analysis of the dataset collected by Trompenaars (1993) originally yielded a seven-factor model; however a subsequent analysis by Smith et al. (1996) identified only three nation-level dimensions. Similarly, a recent publication by Hanges and Dickson (2006) explored the factor structure of the GLOBE data at different levels and found that the final list of dimensions replicates only with the national and organization level data, but not with the data representing individual responses.

Third is the issue of item wording. The level of measurement must be considered not only from the mathematical, but also from the linguistic point-of-view. For example, a question about team-based rewards can inquire about the respondent's personal preferences for reward structure, or whether the respondent believes that people in a given organization or society tend to favor team-based over individual rewards. The review of related literature, including a recent debate in the *Journal of International Business Studies* (Hofstede, 2006; Javidan et al., 2006), reveals that there is no common agreement on the issue. Some scholars point out the inappropriateness of asking questions about individual beliefs and experiences to measure national culture and argue that survey items should be framed at the target level and inquire not about respondents' own values but about the general values of people in their countries (Javidan et al., 2006; Morgeson and Hofmann, 1999). Others question the ability of individuals to provide valid assessments of group-level phenomena, be it their own society or organization or others, and point out that such responses will be largely stereotypical and affected by the respondent's subjective point of reference (Heine et al., 2002; Hofstede, 2006; Smith, 2006).

Our review of the culture survey instruments shows a frequent lack of correspondence between item wording and the intended level of measurement. For example, while Hofstede (1980) focuses on the national level of measurement, virtually all items in his instrument refer to individual preferences and experiences. Furthermore, while items in some instruments clearly differentiate between individual and societal levels (House et al., 2004), others do not. For example, items referring to both individual and societal practices are mixed along a single dimension in the individualism measure developed by Hui (1988): "I would not share my ideas and newly acquired knowledge with my parents" (individual level) and "These days, parents are too stringent with their kids, stunting the development of initiative" (societal level).

In addition to establishing measure equivalence across levels, a cross-cultural scholar must verify that the measure is functionally equivalent across cultures. Because of the trans-border nature of cross-cultural research, it is almost certain that the instrument will have to be translated into different languages. There are a number of challenges associated with item translation and the linguistic aspects seem to be the least problematic. It has long been recognized that a simple word-for-word translation is not adequate, especially if the item wording involves idioms, proverbs, or slang. Simple effective rules have been offered to deal with the purely linguistic challenges of translation (Bontempo, 1993; Laroche et al., 1998), but the problem of item equivalence is a much deeper one and the traditional back and forth translation may not be sufficient to effectively address this challenge.

Two types of cross-culture measurement equivalence issues have been recognized. First is the equivalence of scaling metrics, a.k.a. calibration equivalence (Riordan and Vandenberg, 1994; Ryan et al., 1999; Smith, 2004). This issue refers to whether respondents perceive scale anchors and intervals in the same manner when responding to items in different languages. For

example, a "four" representing "agree" on a five-point Likert scale may be perceived as referring to quite a strong agreement in one language, whereas in another language it may be interpreted as only marginal agreement. The interpretation discrepancy is likely to increase as the scale range increases. This is essentially the issue of intercept differences and can be corrected with techniques similar to those used to correct for extreme response style bias (Cheung and Rensvold, 2000).

Second is the equivalence of factor structure and instrument properties across languages (Ghorpade et al., 1999; van de Vijver and Leung, 1997). This implies that the factor structure of a measure and the instrument properties, such as reliability, must be equivalent across languages. Factor analysis and Structural Equation Modeling could be employed to confirm model structure equivalence across languages (for examples see Marsh, 2003; van de Vijver and Poortinga, 2002). In addition, Item Response Theory (IRT) can be used to further verify item functioning equivalence in different languages (for examples see Bontempo, 1993; Hulin et al., 1982).

Only 30.5% of the studies reviewed indicated that the model structure and instrument property equivalence had been verified across respondent groups that used different languages. However, the trend is quite promising. While the issue had been virtually ignored in the studies of the 80's, more than half of the reports published in the most recent decade indicated that the cross-language factor structure and instrument property equivalence had been analyzed.

5.1. Cross-level and cross-culture equivalence: challenges and best practices

As discussed above, effective methods for verifying measure equivalence across levels of culture and for correcting for possible discrepancies have been offered, many of them cannot be used in small projects as the methods require a large number of groups/societies to be included in the sample. For example, to explore factor structure or measure reliability equivalence at the individual and aggregate data levels, the number of countries should be large enough to meet the sample size requirements for factor analysis techniques. There have been only a few large-scale culture measurement projects that collected data from more than a few dozen countries. On the other hand, establishing measure functioning equivalence across cultures, at least for the individual level of measurement, does not necessarily require that the data represent more than two groups and should be tested in every cross-culture measurement study.

Testing measure properties across levels and cultures is essential for creating an accurate and robust instrument for quantifying culture. First of all, linguistic equivalence and item functioning equivalence in different languages must be established. To minimize construct contamination, items representing a single construct must be worded to refer to a single level of measurement. In addition, we strongly advise that tests are conducted to check for appropriateness of data aggregation to higher levels of measurement, for equivalence of factor structure and instrument properties across levels of measurement, and scale calibration consistency across cultures. If inconsistencies are found, appropriate correction techniques must be used to ensure valid comparison.

6. The reliability and validity of culture measures

Reliability data has been reported for the vast majority of the reviewed culture survey instruments. Generally, reliabilities are assessed separately for each dimension included in the model. Based on the reports provided by the authors of the instruments, the overall average internal consistency reliability, as measured by Cronbach's alpha, was 0.72, ranging from 0.41 to 0.94. The number, however, is somewhat lower based on the reports of the scholars who subsequently used the instruments in later research. For example, Taras and Steel (2006b) meta-analyzed data from 508 studies that used most of the culture survey instruments discussed in this paper. Internal reliability was reported for 63.8% of the studies. Average Cronbach's alpha was 0.67, ranging from 0.41 to 0.82. The internal reliability of Hofstede's VSM, the most popular instrument for quantifying culture, was moderate, averaging 0.68 across the original four factors.

Unfortunately, Cronbach's alpha is usually the only reliability coefficient reported for culture measurement instruments. Test–retest reliability was reported only in twelve papers presenting the 121 reviewed instruments representing roughly 10% of the reviewed measures, averaging 0.58 and ranging from 0.35 to 0.76. Split-half reliability was reported only for six, mainly earlier culture measurement instruments (average 0.69).

Based on the information provided in the studies that introduced reviewed culture measurement instruments, measure validity assessment was conducted in less than 30% of the cases, usually in the larger projects. Item generation and selection was primarily guided by face validity, as evaluated by the authors of the instruments and sometime by external experts. Once the data were collected, the validity of the measure was usually evaluated using criterion-related and convergent validity indicators. Criterion-related validity has typically been evaluated by analyzing correlations between the test scores and a set of theoretically-relevant external variables; for example measures of wealth, freedom, or climate and geography (e.g., Hofstede, 1980; House et al., 2004). Convergent validity has been assessed in most larger culture measurement studies by analyzing the relationship between the obtained national culture scores and findings from earlier instruments, normally those from Hofstede's (1980) and other large-scale projects. Other methods of validity assessment have rarely been used.

6.1. Reliability and validity of culture measures: challenges and best practices

Although reported Cronbach's alphas of the culture measures are generally around and above 0.70, which is commonly accepted as the cut-off point for sufficient internal reliability of a survey instrument, Lance et al. (2006) point out that this criterion

is largely based on an unsubstantiated myth and poor referencing. The work of Nunnally (1978) that is usually cited in relation to the 0.70 limit actually stated that "in many applied settings a reliability of 0.80 is not nearly high enough" and "in those applied settings where important decisions are made with respect to specific test scores, a reliability of 0.90 is the minimum that should be tolerated, and a reliability of 0.95 should be considered the desirable standard" (p. 246). Given that cross-cultural research does not seem to deal with the issues of life and death, a reliability of 0.95 may be too high of a standard, but on the other hand, an observed average Cronbach's alpha of 0.69 seems rather low.

The suspiciously low internal consistencies of many of the reviewed culture measurement instruments may be related to the challenges of culture definition. Essentially, poor internal consistency suggests that the items are measuring multiple unrelated or marginally related constructs. As discussed earlier, construct contamination seems to be a problem of many of the reviewed measures, with values and practices or seemingly unrelated types of values lumped into a single dimension. Commonly-observed mixing of items referring to individual and societal levels and items that seem to relate to constructs other than culture, for example personality, further elevates the problem. It is strongly advised that cleaner, more refined and specific measures of culture are sought to improve measure reliability and precision.

We also advocate for a more comprehensive assessment for reliability and validity. Given that culture by definition is shared and stable, measures of inter-rater (respondent) agreement and temporal stability of the measure are essential to establish that what is measured is indeed a shared and stable construct. A sole measure of internal reliability is certainly not sufficient and a more rigorous assessment is strongly recommended. Similarly, culture survey instrument validation should be done comprehensively including an assessment of face, content, construct validity, as well as validation against various criteria, such as indices from other cross-cultural comparison studies and relevant external constructs such as economic, political, demographic, and if possible behavioral indicators. If possible, the reliability and validity assessment should be done separately at each individual, group, and country level of analysis, unless the scope of the research project and the dataset does not permit for doing so.

7. Discussion and directions for future research

One of the most striking conclusions of our review is that the existing measures of culture are fairly consistent in terms of their approach and closely resemble the methodology used by Hofstede (1980). The vast majority of the existing instruments operationalize culture through values and the data are collected via self-report questionnaires. Generally, scholars attempt to use matched samples and ensure cross-language questionnaire equivalence by double-translating and comparing the different versions. The resulting cultural indices are normally reported as group mean scores along a number of dimensions and findings are validated by correlating the resulting indices with those obtained in other cross-cultural comparison studies or assessing their relationship with external theoretically relevant constructs.

We have witnessed some progress in the way culture has been conceptualized and measured. However, our review indicates that, unfortunately, the progress in measuring culture is mainly made by adding new dimensions to existing models, improving psychometric properties of the questionnaires, and offering new presumably refined sets of cultural indices and rankings. More notable developments included attempts to go beyond studying culture exclusively as values as well as to refine the measurement of cultural dimensions by further narrowing measures to specific facets of culture.

We suggest that future improvements are made in the following areas. Depending on the purpose and scope of the study, the definition of culture and operationalization of its specific dimensions should be more focused to avoid construct contamination. Additional checks should be made to ensure that the used measures are indeed cultural and do not measure related yet conceptually different constructs, such as personality. Next, in an attempt to ensure sample compatibility, it is import to keep in mind sample representativeness, as using matched samples may limit generalizability of findings and thus applicability of the results. While rating items are the most convenient option for cross-cultural surveys, scholars should remember that scenario items generally have higher validity. Furthermore, to improve validity of the findings, it is strongly recommended to consider experimental and behavior frequency data as an alternative to self-response questionnaires. Regardless of the type of data used, it is also recommended to analyze not only group means, but also other measures of data distribution, including variance, skewness, and kurtosis. Scholars should also remember that simple double translation is not sufficient to ensure cross-cultural equivalence of survey instruments. Equivalence of factor structures, reliability and other psychometric properties, and item functioning for versions of the survey in different languages should also be established. Additionally, surveys should be designed and scored so that cross-cultural differences in response styles, such as the propensity to extreme or moderate responses and acquiescence bias, are controlled for.

These, however, would be marginal improvements. The following important fundamental questions of culture conceptualization and measurement should be considered to achieve any considerable progress:

Are cultures values and are values culture? Virtually all reviewed instruments are based on the assessment of values that are presumably determined by culture. However, it is commonly agreed that culture is a multi-layer construct represented by observable artifacts and practices at the surface and by tacit attitudes, values, and basic assumptions at the core (e.g., Hofstede, 2001; Trompenaars, 1993). Is solely measuring cultural values adequately capturing the phenomenon of culture? So far, no conclusive empirical results are available to describe the nature and magnitude of the relationship between different layers of culture and it remains uncertain whether culture measures are generalizable across levels of the construct. On the other hand, with the increase in popularity of value-based measures of culture and the growing evidence of value differences between cultural entities, it has become common to assume that all types of values are cultural. However, the factors that shape values remain largely unexplored and it is plausible to suggest that many types of values and attitudes are *not* cultural.

Are we actually measuring what we think we are measuring? It is questionable that the degree to which all constructs measured by the reviewed culture survey instruments actually relate to culture, and not, for example, to personality. As discussed earlier, numerous existing definitions of culture are worded differently and emphasize different elements, but most of them agree that culture is remarkably stable, culture is shared by individuals in a cultural entity, as well as that culture is more than just values or beliefs.

In regard to the first assumption, in an attempt to highlight the unchangeable nature of culture, cultural values have even been compared to "software of mind" and "hardwired mental programs" (Hofstede, 1991). In other words, in theory, cultural values should represent a permanent trait, not a temporal state of an individual or an ethnos. However, a number of studies have found that what we call "national cultures" has been changing quite rapidly (e.g., Fernandez et al., 1997; Inglehart and Welzel, 2005; Taras and Steel, 2006b). Similarly, rather low test–retest reliability coefficients suggest that the constructs measured by many of the reviewed instruments may actually be temporal states and perhaps such terms as "currently prevailing beliefs" or "currently preferred practices" would be more appropriate to describe the measured constructs. Unfortunately, test–retest reliability has been reported only for a very few instruments and any conclusions would be premature at this time. To address the issue, we recommend that in addition to evaluating basic psychometric properties of the instrument such as internal reliability and factor loadings, high temporal stability of the measured constructs be established, for example, by the means of test–retest reliability analysis.

Furthermore, because cultures are seen as shared traits, the test scores must effectively predict membership in culture-specific groups and effectively differentiate between members of different cultural entities. A simple test of significance of the difference between group means is not sufficient to conclude that differences between the entities are indeed cultural. A more comprehensive analysis of the cross-group discriminant validity of the measure must be performed before it can be labeled "cultural" (Au and Cheung, 2004; Durvasula et al., 2006).

How well are self-report questionnaires measuring culture? Self-report questionnaires have been virtually the only way to collect data to quantify culture. Unfortunately, the degree to which this approach can provide valid data to measure the phenomenon remains questionable. First of all, there may be a substantial difference between the perceptions of the respondents and reality. An average person simply has limited abilities to provide exact numerical answers to questions of such types as "How do you feel about…?", "How important …?" and "To what extent do you believe…?" Even if the survey item inquires about such identifiable things as one's individual past or likely future behaviors, it is doubtful that one will be able to provide perfectly accurate answers. However, even if one could, language of the survey, differences in response styles, and differences in interpretation of the scale anchors are all threats to validity of the data obtained through self-report, as discussed earlier. Several methods, including combining positively- and negatively-worded items (McCrae and Costa, 1997) and various within-subject standardization procedures (Hofstede, 1980; Leung and Bond, 1989) have been suggested to correct for systemic response biases in intercultural surveys, but they are not without limitations (Smith, 2004). Future research must assess these concerns, determine what is negligible in impact and what is not, and then seek ways to reduce any significant sources of contamination.

Are national/regional culture averages becoming obsolete? Years ago, nationality was probably an acceptable proxy for culture. In some cases, regional or ethnic differentiation had to be made, but generally, one's area of residence and ethnic background may have been good predictors of one's culture. In today's "global village," geographical boundaries are becoming less relevant. While people across the world have different cultural values, they are influenced by more than just geography. There may be a greater variation in cultural values across generations, professions, interest communities, or socio-economic classes, than across countries. Future research should re-examine the boundaries of cultural clusters and devote more attention to the individual level of analysis.

How do we address acculturation and situation-specific cultural accommodation? With rapidly increasing migration across borders, country of origin becomes less relevant in defining one's cultural values. Unfortunately, very little is known about change in one's cultural values as one is exposed to a new cultural environment. Are cultural values really hardwired, or do they change following, for example, immigration? Moreover, what are the cultural values of those who grow up in bi-cultural families or those who learn one culture from their immigrant parents, but are faced with the culture of their host society outside the home? Can people change their values and behaviors depending on cultural circumstances? Answers to these basic questions are still to be found.

What are the alternatives to the quantitative approach to studying culture? Finally, the qualitative approach has been successfully used for centuries and remains the main mode of enquiry in anthropology, the original field of cross-cultural studies. With the increasing popularity of the quantitative approach to studying any social phenomenon, including culture, are we limiting the richness of the data we work with? Quantitative approaches are useful for making comparisons across cultures but often fail to capture the unique variance that is specific to a few or only one culture alone.

As noted by Littauer (1960), "most people, and almost all scientists, would say that measurement and science were synonymous" (p. 36). Our ability to measure culture, quantitatively and qualitatively, is critical to our progress in cross-cultural studies. To this end, it is important to offer congratulations to the many researchers who have made substantive progress in cultural assessment and consequently to culture's science. Nevertheless, this praise must be tempered. Despite our advances, this review indicates that our ability to measure culture still remains incomplete. To keep momentum, we should not succumb to the hubris of believing we have already brought the field to its logical conclusion. The science of culture can only advance by acknowledging our strengths and weaknesses, allowing for further refinement of existing instruments and techniques as well as in the development of alternative approaches beyond assessing values with self-report questionnaires. By providing a comprehensive review of the issues, challenges, and best practices in the area of culture measurement and by identifying major trends and gaps, the present study builds a foundation for further development of culture measurement techniques.

References

Allik, J., McCrae, R.R., 2004. Toward a geography of personality traits: patterns of profiles across 36 cultures. Journal of Cross-Cultural Psychology 35 (1), 13–28.

Au, K.Y., Cheung, M.W.L., 2004. Intra-cultural variation and job autonomy in 42 countries. Organization Studies 25 (8), 1139–1362.

Benet-Martinez, V. (2000). Examining interplay between personality and culture: Personality correlates of cultural, ethnic, and linguistic orientations Unpublished manuscript.

Benet-Martinez, V., Karakitapoglu-Aygun, Z., 2003. The interplay of cultural syndromes and personality in predicting life satisfaction: comparing Asian Americans and European Americans. Journal of Cross-Cultural Psychology 34 (1), 38–61.

Bennett, M., 1977. Response characteristics of bilingual managers to organizational questionnaires. Personnel Psychology 30 (1), 29–36.

Bing, J.W., 2005. The use and misuse of questionnaires in intercultural training. Retrieved Oct, 10, 2005, from http://www.itapintl.com/useandmisuse.htm.

Bond, M.H., Leung, K., Au, A., et al., 2004. Culture-level dimensions of social axioms and their correlates across 41 cultures. Journal of Cross-Cultural Psychology 35 (5), 548–570.

Bontempo, R., 1993. Translation fidelity of psychological scales: an item response theory analysis of an individualism–collectivism scale. Journal of Cross-Cultural Psychology 24, 149–166.

Carver, C., Scheier, M., Weintraub, J.K., 1989. Assessing coping strategies: a theoretically based approach. Journal of Personality and Social Psychology 56 (2), 267–283.

Cheung, G.W., Rensvold, R.B., 2000. Assessing extreme and acquiescent response sets in cross-cultural research using structural equations modeling. Journal of Cross-Cultural Psychology 31 (2), 187–212.

Chew, K. H. (1996). Beyond individualism–collectivism: additional constructs to consider. Unpublished doctoral dissertation, University of Oregon, Eugene, WA, US.

Cook, T.D., Campbell, D.T., 1979. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Houghton Miffin Company, Boston.

Costa Jr., P.T., McCrae, R.R., 1992. NEO PI-R: Professional Manual. Psychological Assessment Resources, Odessa, FL.

Denny, R.M., Sunderland, P.L., 2005. Researching cultural metaphors in action: metaphors of computing technology in contemporary U.S. life. Journal of Business Research 58 (10), 1456–1477.

Dorfman, P., Howell, J.P., 1988. Dimensions of national culture and effective leadership patterns: Hofstede revisited. In: Farmer, R.N., McGoun, E.G. (Eds.), *Advances in International Comparative Management*. JAI Press, London, UK, p. 172-150.

Durvasula, S., Netemeyer, R.G., Andrews, J.C., et al., 2006. Examining the cross-national applicability of multi-item, multi-dimensional measures using generalizability theory. Journal of International Business Studies 37 (4), 469–481.

Earley, P.C., 1993. East meets West meets Mideast: further explorations of collectivistic and individualistic work groups. Academy of Management Journal 36 (2), 319–348.

England, G.W., 1967. Personal value systems of American managers. Academy of Management Journal 10, 53–68.

Eylon, D., Au, K.Y., 1999. Exploring empowerment: cross-cultural differences along the power distance dimension. International Journal of Intercultural Relations 23 (3), 373–385.

Fernandez, D.R., Carlson, D.S., Stepina, L.P., et al., 1997. Hofstede's country classification 25 years later. Journal of Social Psychology 137 (1), 43–54.

Furnham, A., 1984. Personality and values. Personality and Individual Differences 5 (4), 483–485.

Furrer, O., Liu, B.S.-C., Sudharshan, D., 2000. The relationships between culture and service quality perceptions: basis for cross-cultural market segmentation and resource allocation. Journal of Service Research 2 (4), 355–371.

Gannon, M., 2004. Understanding Global Cultures: Metaphorical Journeys Through 28 Nations, Clusters, and Continents, 3 ed. Sage, Thousand Oaks, CA.

Gannon, M., Locke, E.A., Gupta, A., et al., 2005–2006. Cultural metaphors as frames of reference for nations. International Studies of Management & Organization 35 (4), 37–47.

Geertz, C., 1973. The Interpretation of Culture. Basic Books, New York.

Gelfand, M., Nishii, L.H., Chan, D.K.S., et al., 1998. Toward a theory of tightness–looseness: further empirical evidence from the U.S. and Japan. Paper presented at the Annuan Conference of the International Association of Cross-Cultural Psychology, Bellingham, WA.

Gelfand, M.J., Nishii, L.H., Raver, J.L., 2006. On the nature and importance of cultural tightness–looseness. Journal of Applied Psychology 91 (6), 1225–1244.

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

Haire, M., Ghiselli, E.E., Porter, L.W., 1966. Managerial Thinking: An International Study. Wiley, New York.

Hall, E.T., 1977. Beyond Culture. Doubleday, New York.

Hanges, P.J., Dickson, M.W., 2006. Agitation over aggregation: clarifying the development of and the nature of the GLOBE scales. Leadership Quarterly 10 (5), 522–536.

Harzing, A.-W., 2006. Response styles in cross-national survey research. International Journal of Cross Cultural Management 6 (2), 243–266.

Harzing, A.-W., Maznevski, M.L., 2002. The interaction between language and culture: a test of the cultural accommodation hypothesis in seven countries. Language and Intercultural Communication 2 (2), 120–139.

Heine, S.J., Lehman, D.R., Peng, K., et al., 2002. What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. Journal of Personality and Social Psychology 82 (6), 903–918.

Hofstede, G., 1980. Culture's Consequences: International Differences in Work-Related Values. Sage Publications, Beverly Hills, CA.

Hofstede, G., 1991. Cultures and Organizations: Software of Mind. McGraw Hill, London.

Hofstede, G., 2001. Culture's Consequences: Comparing Values, Behaviors, Institutions, and Values Across Nations, 2 ed. Sage Publications, London.

Hofstede, G., 2002a. Dimensions do not exist: a reply to Brendan McSweeney. Human Relations 55 (11), 11355–11361.

Hofstede, G., 2002b. The pitfalls of cross-national survey research: a reply to the article by Spector et al. on the psychometric properties of the Hofstede Values Survey Module 1994. Applied Psychology 51 (1), 170–173.

Hofstede, G., 2006. What did GLOBE really measure? Researchers' minds versus respondents' minds. Journal of International Business Studies 37 (6), 882–897.

Hofstede, G., McCrae, R.R., 2004. Personality and culture revisited: linking traits and dimensions of culture. Cross-Cultural Research 38 (1), 52–89.

House, R.J., Hanges, P.J., Javidan, M., et al. (Eds.), 2004. Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies. Sage Publications, Thousand Oaks, CA.

Hui, C.H., 1988. Measurement of individualism–collectivism. Journal of Research in Personality 22, 17–36.

Hui, C.H., Triandis, H.C., 1986. Individualism–collectivism: a study of cross-cultural researchers. Journal of Cross-Cultural Psychology 17 (2), 225–248.

Hui, C.H., Triandis, H.C., 1989. Effects of culture and response format on extreme response style. Journal of Cross-Cultural Psychology 20 (3), 296–309.

Hui, C.H., Yee, C., 1994. The shortened individualism–collectivism scale: its relationship to demographic and work-related variables. Journal of Research in Personality 28 (4), 409–424.

Hulin, C.L., Drasgow, F., Komocar, J., 1982. Applications of item response theory to analysis of attitude scale translations. Journal of Applied Psychology 67 (6), 818–826.

Inglehart, R., 1997. Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies. Princeton University Press, Princeton, NJ.

Inglehart, R., Basanez, M., Moreno, A., 2004. Human Values and Beliefs: A Cross-Cultural Sourcebook. Ciglo Veintiuno Editores, Delegacion Coyoacan, Mexico.

Inglehart, R., Welzel, C., 2005. Modernization, Cultural Change and Democracy: The Human Development Sequence. Cambridge University Press, New York.

Javeline, D., 1999. Response effects in polite cultures: a test of acquiescence in Kazakhstan. Public Opinion Quarterly 63, 1–28.

Javidan, M., House, R.J., Dorfman, P.W., et al., 2006. Conceptualizing and measuring cultures and their consequences: a comparative review of GLOBE's and Hofstede's approaches. Journal of International Business Studies 37 (6), 897–914.

Kim, J.-O., Mueller, C.W., 1978. Factor Analysis: Statistical Methods and Practical Issues. Sage Publications, Beverly Hills.

Kluchhohn, F.R., Strodtbeck, F.L., 1961. Variations in Value Orientations. Row, Peterson, Evanston, IL.

Koltko-Rivera, M.E., 2004. The psychology of worldviews. Review of General Psychology 8 (1), 3–58.

Kroeber, A.L., Kluckhohn, C., 1952. Culture: A Critical Review of Concepts and Definitions. Vintage Books, New York.

Kuhn, M.H., McPartland, R., 1954. An empirical investigation of self attitudes. American Sociological Review 19, 68–76.

Lance, C.E., Butts, M.M., Michaels, L.C., 2006. The sources of four commonly reported cutoff criteria: what did they really say? Organizational Research Methods 9 (2), 202–220.

Laroche, M., Kim, C., Tomiuk, M.A., 1998. Translation fidelity: an IRT analysis of Likert-type scale items from a culture change measure for Italian-Canadians. Advances in Consumer Research 25, 240–245.

Leung, K., Bond, M.H., 1989. On the empirical identification of dimensions for cross-cultural comparisons. Journal of Cross-Cultural Psychology 20 (2), 133–151.

Leung, K., Bond, M.H., 2004. Social axioms: a model for social beliefs in multicultural perspective. In: Zanna, M.P. (Ed.), Advances in Experimental Social Psychology, Vol. 36. Elsevier Academic Press, San Diego, CA, pp. 119–197.

Leung, K., Bond, M.H., de Carrasquel, S.R., et al., 2002. Social axioms: the search for universal dimensions of general beliefs about how the world functions. Journal of Cross-Cultural Psychology 33 (3), 286–302.

Littauer, S.B., 1960. On some aspects of technological measurement. Management Technology 1 (1), 36–45.

Marin, G., Gamba, R.J., Marin, B.V., 1992. Extreme response style and acquiescence among Hispanics. Journal of Cross-Cultural Psychology 23 (4), 498–509.

Marsh, D. R. (2003). Establishing the cross-cultural measurement equivalence of individualism–collectivism. Unpublished doctoral dissertation, Washington State University.

Maznevski, M.L., DiStefano, J.J., 1995. Measuring Culture in International Management: the Cultural Perspectives Questionnaire. The University of Western Ontario Working Paper Series, p. 95-39.

McCrae, R.R., Costa Jr., P.T., 1997. Personality trait structure as a human universal. American Psychologist 52 (5), 509–516.

McSweeney, B., 2002. Hofstede's model of national cultural differences and their consequences: a triumph of faith — a failure of analysis. Human Relations 55 (1), 89–118.

Miller, J.G., 2002. Bringing culture to basic psychological theory—beyond individualism and collectivism: comment on Oyserman et al. (2002). Psychological Bulletin 128 (1), 97–109.

Morgeson, F.P., Hofmann, D.A., 1999. The structure and function of collective constructs: implications for multilevel research and theory development. The Academy of Management Review 24 (2), 249–265.

Nunnally, J.C., 1978. Psychometric Theory, 2 ed. McGraw-Hill, New York.

Offermann, L.R., Hellmann, P.S., 1997. Culture's consequences for leadership behavior: national values in action. Journal of Cross-Cultural Psychology 28 (3), 342–351.

Oyserman, D., Coon, H.M., Kemmelmeier, M., 2002. Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. Psychological Bulletin 128 (1), 3–72.

Paulson, S.K., 2005. Teaching international business concepts through the exchange of cultural metaphors. Journal of Teaching in International Business 16 (4), 81–99.

Peng, K., Nisbett, R.E., Wong, N.Y., 1997. Validity problems comparing values across cultures and possible solutions. Psychological Methods 2 (4), 329–344.

Peterson, M.F., Castro, S.L., 2006. Measurement metrics at aggregate level of analysis: implications for organizational culture research and the GLOBE project. Leadership Quarterly 17, 506–521.

Riordan, C.M., Vandenberg, R.J., 1994. A central question in cross-cultural research: do employees of different cultures interpret work-related measures in an equivalent manner? Journal of Management 20 (3), 643–671.

Rokeach, M., 1973. The Nature of Human Values. Free Press, New York.

Ryan, A.M., Chan, D., Ployhart, R.E., et al., 1999. Employee attitude surveys in a multinational organization: considering language and culture in assessing measurement equivalence. Personnel Psychology 52, 37–58.

Schaffer, B.S., Riordan, C.M., 2003. A review of cross-cultural methodologies for organizational research: a best-practices approach. Organizational Research Methods 6 (2), 169–215.

Schimmack, U., Oishi, S., Diener, E., 2005. Individualism: a valid and important dimension of cultural differences between nations. Personality and Social Psychology Review 9 (1), 17–31.

Schwartz, S.H., 1992. Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In: Zanna, M.P. (Ed.), Advances in Experimental Social Psychology. Academic Press, San Diego, CA, pp. 1–65.

Schwartz, S.H., 1994. Beyond individualism/collectivism: new cultural dimensions of values. In: Kim, U., Triandis, H.C., Kagitcibasi, C., Choi, S.C., Yoon, G. (Eds.), Individualism and Collectivism: Theory, Methods and Applications. Sage Publications, London, pp. 85–119.

Schwartz, S.H., Boehnke, K., 2004. Evaluating the structure of human values with confirmatory factor analysis. Journal of Research in Personality 38 (3), 230–255.

Schwarz, N., 1999. Self-reports: how the questions shape the answers. American Psychologist 54 (2), 93–105.

Singelis, T.M., 1994. The measurement of independent and interdependent self-construals. Personality & Social Psychology Bulletin 20 (5), 580–591.

Singelis, T.M., Triandis, H.C., Bhawuk, D.P.S., et al., 1995. Horizontal and vertical dimensions of individualism and collectivism: a theoretical and measurement refinement. Cross-Cultural Research: the Journal of Comparative Social Science 29 (3), 240–275.

Smith, P.B., 2004. In search of acquiescent response bias. Journal of Cross-Cultural Psychology 35 (1), 50–61.

Smith, P.B., 2006. When elephants fight, the grass gets trampled: the GLOBE and Hofstede projects. Journal of International Business Studies 37 (6), 915–922.

Smith, P.B., Dugan, S., Trompenaars, F., 1996. National culture and the values of organizational employees: a dimensional analysis across 43 nations. Journal of Cross-Cultural Psychology 27 (2), 231–264.

Soudijn, K.A., Hutschemaekers, G.J.M., van de Vijver, F.J., 1990. Culture conceptualisations. In: van de Vijver, F.J.R., Hutschemaekers, G.J.M. (Eds.), The investigation of culture: current issues in cultural psychology. Tilburg University Press, Tilburg, The Netherlands, p. 209.

Spector, P.E., 1988. Development of the work locus of control scale. Journal of Occupational Psychology 61 (4), 335–340.

Spector, P.E., Cooper, C.L., Sparks, K., 2001. An international study of the psychometric properties of the Hofstede Values Survey Module 1994: a comparison of individual and country/province level results. Applied Psychology: An International Review 50 (2), 269–281.

Stedham, Y.E., Yamamura, J.H., 2004. Measuring national culture: does gender matter? Women in Management Review 19 (5), 233–243.

Steel, P., Ones, D.S., 2002. Personality and happiness: a national-level analysis. Journal of Personality & Social Psychology 83 (3), 767–781.

Taras, V., 2008. Culture Survey Catalogue: Original Items, Scoring Keys and Psychometric Properties of 134 Instruments for Measuring Cultural Values and Behaviors. from http://ucalgary.ca/~taras/_private/Culture_Survey_Catalogue.pdf.

Taras, V., Steel, P., 2005. Cross-cultural differences and dynamics of cultures over time: a meta-analysis of Hofstede's taxonomy. Paper presented at the Academy of Management conference, Honolulu, HI, USA.

Taras, V., Steel, P., 2006b. Culture as a consequence: a multilevel multivariate meta-analysis of the effects of individual and country characteristics on work-related cultural values. Best Papers Proceedings, the Academy of Management Annual Meeting, Atlanta, GA.

Taras, V., Steel, P., 2006a. Improving cultural indices and rankings based on a meta-analysis of Hofstede's taxonomy. Paper presented at the Academy of International Business Annual Meeting, Beijing, China.

Thorndike, E., 1939. On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. American Journal of Psychology 52 (1), 122–124.

Triandis, H.C., 1994a. Cross-cultural industrial and organizational psychology, 2 ed. : Handbook of Industrial and Organizational Psychology, vol. 4. Consulting Psychologists Press, Inc, Palo Alto, CA, US, pp. 103–172.

Triandis, H.C., 1994b. INDCOL. Unpublished Research Scale on Individualism and Collectivism. University of Illinois, Champaign.

Triandis, H.C., Bontempo, R., Betancourt, H., et al., 1986. The measurement of the ethic aspects of individualism and collectivism across cultures. Australian Journal of Psychology 38 (3), 257–267.

Triandis, H.C., Chen, X.P., Chan, D.K., 1998. Scenarios for the measurement of collectivism and individualism. Journal of Cross-Cultural Psychology 29 (2), 275–289.

Triandis, H.C., McCusker, C., Betancourt, H., et al., 1993. An etic–emic analysis of individualism and collectivism. Journal of Cross-Cultural Psychology 24 (3), 366–383.

Trompenaars, F., 1993. Riding the Waves of Culture: Understanding Diversity in Global Business. Irwin Professional Publishing, Chicago, IL.
Trubisky, P., Ting-Toomey, S., Lin, S.-L., 1991. The influence of individualism–collectivism and self-monitoring on conflict styles. International Journal of Intercultural Relations 15, 65–84.
van de Vijver, F.J., Leung, K., 1997. Methods and data analysis of comparative research, In: Berry, J.W., Poortinga, Y.H., Pandey, J. (Eds.), 2 ed. Handbook of Cross-Cultural Psychology: Theory and Methods, Vol. 1. Allyn & Becon, Boston, MA, pp. 406–432.
van de Vijver, F.J., Poortinga, Y.H., 2002. Structural equivalence in multilevel research. Journal of Cross-Cultural Psychology 33 (2), 141–156.
Vandello, J.A., Cohen, D., 1999. Patterns of individualism and collectivism across the United States. Journal of Personality and Social Psychology 77 (2), 279–292.
Wagner, J.A., Moch, M.K., 1986. Individualism–collectivism: concept and measure. Group & Organization Studies 11 (3), 280–303.