Measurement and Validation of Hall's Construct of Context:

Extending the Cross-Cultural Measurement Paradigm

Abstract:

This study develops a scale to measure the communication construct of Hall's (1976) context. The authors validate and test the scale against two other competing versions with data from 23 countries. They develop a new type of scale conducted with bipolar nouns, instead of traditional adjectives, that they term "semantic differential metaphorical." The results suggest that a new semantic differential metaphorical scale outperforms both Likert and traditional semantic differential scales when measuring context and perhaps other cultural communication constructs. The authors develop and compare the scales on the basis of traditional measures (reliability, invariance, and nomological validity) and examine them for the propensity to be affected by individual response style biases.

Keywords: context, cross-cultural communication, scale development, Hall

Measure what is measurable, and make measurable what is not so.

—Galileo (1564–1642)

Context is pervasive throughout marketing communications and is a key component of culture. As a topic in intercultural business communication, high and low context has appeared for decades in virtually every text on international business and international marketing. Originally presented by Hall (1976) in his book *Beyond Culture*, the model is often discussed in the literature, although efforts to validate a measure of context have been unsuccessful (Cardon 2008). The construct of context remains anecdotal, ill defined, and consequently of limited use in international marketing strategy and research except at the most conceptual level.

Context has been used to study many marketing phenomena such as an antecedent of persuasion (Aaker and Maheswaran 1997), management of information processing (Luna, Ringberg, and Peracchio 2008), elaboration or interpretation of messages (Kroll and De Groot 1997), evaluation of brand extensions (Monga and John 2007), choice of communication media (Richardson and Smith 2007), and pricing (Nguyen, Heeler, and Zinaida 2007). For the most part, that research has simply assumed that Hall's (1976) ranking is accurate and has compared countries using those rankings.

We set out to define and develop a scale to measure context. We test a new approach to survey design using a metaphorical approach that appears to be superior to traditional semantic differential and Likert scales when measuring constructs related to cross-cultural communications such as context.

DEFINING CONTEXT

The American Heritage Dictionary defines context as "the part of a written or spoken statement in which a word or passage at issue occurs and that often specifies its meaning." Hall (1976) defined high-context communication to be when there is little information contained in the explicit part of the message but relatively more information sent indirectly that requires a great deal of cultural programming to decipher. The physical context in which the message is delivered or the person's demeanor determines the true meaning of the message. Low-context messages are readily discernible from the explicit portion of the message and do not demand training in cultural nuance to understand what is being communicated.

A review of global/international marketing text books reveals that there is broad agreement on the rankings of certain countries and characteristics of high- versus low-context communication and countries. Most of these can be traced either to Hall's (1976) original rankings or to Hall and Hall (1987; 1990) none of whom actually measured context. Unfortunately, discussions of context mostly rely on either descriptors associated with context or a comparison of communication styles in low- versus high-context cultures. When descriptions are stated, they are typically provided separately for high context (explicit communication) and low context (message relies on environment). The major, uniform element of context is the degree to which the presenter precisely states his or her position. Thus, context can be captured simply as follows: Context represents the degree of explicitness of written/verbal communication within a culture.

Whether dichotomous (e.g. east vs. west) or continuous, countries are ranked or positioned purely on the basis of face validity. Such approaches are insufficient because of a self-referent bias (Cardon 2008). These methods satisfy neither the demand for rigor in academic research nor the need for precision in applied marketing research. For example, there is often a tendency to treat Asia as monolithic. Yet even among similar East Asian cultures, such as China, Japan, and Korea, there are differences in context. Moreover, Asia consists of far more diverse cultures than the East Asian triad, many of which differ dramatically in terms of context.

Similarly, the Muslim world of the Near East differs greatly from that of Indonesia and Malaysia. Thus, lumping context into East and West or high and low is insufficient.

REVIEW OF LITERATURE

Hall (1976) laid the framework of the construct for context but never justified his ordering or empirically tested his assumptions. Subsequent works have encountered great difficulty in developing measures of context that allow researchers to clearly identify variation from one country to another. One of the most rigorous of these efforts was by Gudykunst and colleagues (1996), who contrasted dimensions of high and low contrast. They derived the dimensions empirically and inductively rather than theoretically and inductively from four countries. Kim, Pan, and Park (1998) examined two items (out of five) that focused on communication in their study of China, Korea and the United States. However, no significant difference was found among the countries. Ohashi (2000) pursued the argument that context is a continuum and therefore should be studied using countries that represented extremes from Hall's perspective. She concluded that Japan was high context and the United States low context, even though both were near the midpoint of her seven-point scales. Richardson and Smith (2007) find a similar outcome, in that both the United States and Japan were near the midpoint of the scale rather than showing any degree of polarization.

Only a few studies have followed Hofstede's (2001) recommendation to use enough countries (a minimum of 10 to 15) to develop etic dimensions. Koeszegi, Vetschera, and Kersten (2004) and Trompenaars (1994) both met this standard of large country samples. However, the outcomes were contrary to what Hall's (1976) continuum indicates.

The purpose of the current research is to (1) develop a scale that captures the concept of context and (2) extend the paradigm of international scale development. Although several studies have inductively described context as being multidimensional, this is largely due to measurement issues rather than to the concept itself. Hall's (1976) original concept and that adopted by all global/international marketing textbooks is unidimensional from high to low based on the degree to which meaning is carried in words instead of the environment. Thus, we adopt context as a unidimensional continuum.

The first step in this process was to develop a scale that would function as an etic dimension. Following Churchill's (1979) well-accepted procedures for initial scale development and Douglas and Craig's (2006) adapted etic approach, we develop a scale to measure Hall's (1976) concept of intercultural context. In addition, we develop and compare and contrast several scale types.

SCALE DEVELOPMENT

Initial Scale Development

On the basis of Churchill's (1979) procedure for scale development, we initially identified 27 items using a seven-point Likert-type scale. A pretest of 218 U.S. college students further narrowed the list to 13 items through principle components analysis and reliability analysis. We then pretested the remaining items across three countries with different cultural context levels: the United States (low), China (high), and Belgium (middle). The initial results indicated five factors, low reliabilities on many factors, and nonsimilar loading patterns across countries.

Literal and Comparable Translation

Following Douglas and Craig's (2006) suggestions, we disaggregated the study to review the research problem. A discussion ensured with seven colleagues, each with scale development expertise or training and who represented 15 unique languages and previously residence in more than 23 countries. After the review, several concerns and issues became quickly apparent.

Equivalent translations were nearly impossible for some scale items. Words that had subtle shades of meaning in English were construed as interchangeable in other languages or lacked any reasonable counterpart. This led to the belief that previous studies might have a pseudo tautology built into the scales—for example, using the word "explicit" necessitates that the respondents understand what "explicit" means across a range of contexts rather than having one standardized meaning. The group suggested adjectives that could be translated into several languages.

There were also issues with response styles with Likert-type endpoints ("strongly disagree/strongly agree"). Different cultures tend to have different response styles (Johansson 2003). For example, the Japanese tend to answer toward the middle of scales to avoid standing out (Hult et al. 2008). The collaborators believed that these particular endpoints could aggravate already existing tendencies for differential response styles. In essence, although two cultures could view the term "explicit" somewhat differently, the response is further confused by adding differential cultural influences of tendencies to "agree" or "disagree." In addition, people who are more culturally isolated (speak fewer languages relative to their peers and travel less) tend to view their language use differently than their peers.

In addition, some of the items included measures related to nonverbal communication (e.g., hand gestures, written communication); these items tended to load individually or not at all. The contributors indicated that nonverbal communication may be a different dimension or construct altogether and should be excluded from the study.

Finally, concerns were expressed about the referencing of individual versus culture in items. Of the original items, nine referenced the individual ("I say precisely what I mean") and four referred to the language in general ("Language is an art, not a science"). The results indicated that there was much more heterogeneity in responses for individual- than for language-referent items. In addition, the initial analysis suggested that individual-referent items were subject to greater variation across demographics (specifically gender) and had much poorer principle components analysis loadings and reliability results. The research group suggested that the proper measurement level is individual perceptions of cultural context (rather than either individual or language context level). It avoids the problems of confusing a single language use across cultures (e.g., English in the United States, United Kingdom, and India) and allows the use of a non-native language when necessary (e.g., English in India, Russian in the former Soviet Union).

Scale Types

We tested Likert and semantic differential scales to determine which would give the best results in measuring cultural context. Alreck and Settle (1985) provide an overview of the relative merits and shortcomings of these scaling methods. The Likert scale measures the degree to which a respondent agrees or disagrees with a statement. It is flexible in that a researcher can use few words or very lengthy items, which is important because functional equivalents of a word or phrase rather than a direct translation are important in international research (Kroll and De Groot 1997). Being able to sum the results enables the researcher to measure a more general construct, such as context. This is especially important given the evidence that many cultures tend to answer in the center of the scale rather than toward the extremes (Hult et al. 2008). We

followed Alreck and Settle's suggestions by having multiple items, composing statements typical of a situation common to all respondents, and providing a seven-point scale to minimize neutral responses.

Because we want to define and measure perception of a person's own communication style, we chose four sets of adjectives that were polar opposites in our semantic differential scale. We selected these with an eye toward Hall's (1976) descriptions of the degree of specificity or vagueness and artistic versus task orientations of communication. The goal was to allow respondents to delineate how they communicate and to identify an ideal or pure form of their communication, as they perceived it. It was critical that all pairs of adjectives converge on a single dimension. The advantage of this method over the previous one is that it should eliminate the bias inherent when a respondent does not want to offer his or her opinion in fear that it differs from the larger group (a collectivist outcome). There is no overt agreement contained in this scale as there is in Likert scales.

We also developed an additional scale, which we term "metaphoric semantic differential." We did this to ensure that language differences did not interfere with understanding the meaning of the adjective and what it measured (Douglas and Craig 2006, 2007). We selected nouns (rather than adjectives) that were unlikely to be subject to much individual interpretation or cultural nuance. For example, respondents from any culture should agree a feature of a rock is hardness. Holistic thinking would not change this quality or its description. Its counterpart, a flower, can be universally agreed to be soft and yielding. The expected outcome was to achieve an etic quality that would overcome linguistic and cultural filters that threaten semantic differential endpoints or descriptors. In addition, the use of referent nouns rather than adjectives allows simpler translations into most languages. A review of typical scale books indicated no previous usage of nouns as descriptive endpoints for scaling procedures.

Revised Scales

A revision of the original scales resulted in 16 items using three scale types (see table 2). For comparability, all scales were seven-point scales. In addition, the scales referenced the culture rather than the individual or language. Each was constructed so that it could be translated into multiple languages with minimal problems using an etic framework (Douglas and Craig 2006). Items that performed relatively well in the pretest were examined and revised by multiple people until they were comfortable translating the scale for both literal and symbolic meaning from English into the following languages: Chinese, French, Japanese, German, Russian, Slovenian, Croatian, Spanish, Lithuanian, Dutch, and Turkish. This constitutes at least one language from most of the major language groups. We developed the semantic differential metaphorical (SDM) scale in part from the pretest, but mostly from the input of the research group. Given the size of the survey and Douglas and Nijssen's (2003) suggestion that shortened versions of scales be used in cross-cultural research, we limited the scale items to five per scale except for the SDM, for which we allowed an additional scale item because it had not yet been pretested (see table 1).

Insert table 1 about here

We subjected the scale items to principle components analysis, reliability analysis, and confirmatory factor analysis (CFA) in the United States, Italy, France, and the Philippines on the basis of convenience and cultural differentiation (see table 2).

Insert table 2 about here

The scales were purified by removing items with lower loadings and those that increased the reliability by removal. We removed the first item in the Likert-type and SDM-type scales (SDAExplici and SDM1NATURE) and the last item in the Likert-type scale (AL5Depend). With the exception of one item on the semantic differential adjective (SDA) scale in Italy and one on the Anchored Likert (AL) in France, all loadings were above .50 and loaded on a single component for each scale type.

SCALE COMPARISON AND VALIDATION

Hofstede (2001) suggests that cross-cultural scales should be validated in at least 10 countries. This is further reinforced by Douglas and Craig's (2006) call to achieve "purposive selection" to ensure variance on characteristics of interest. In their study, Hall (1976) and Hall and Hall (1987) arranged the following countries from lowest to highest context: Germany, Scandinavian countries, North American countries, France, United Kingdom, Italy, Latin American countries, Arab countries, China, and Japan. With this list and Hofstede's (2001) measures of cultural difference, we chose 23 countries (see table 3). We chose many of Hall's countries and included a variety of emerging economies because their role in international trade is rapidly expanding in importance and their study in previous research has been relatively sparse (Burgess and Steenkamp 2006).

Insert table 3 about here

We chose college students for the study because of several factors: (1) their relative homogeneity of extraneous influences (Coulter et al. 2005; Strizhakova, Counter, and Price 2008), (2) their relatively high exposure to global commerce (Gidley 2002; Kjeldgaard and Askegaard 2006), and (3) their relatively high exposure to multiple languages and cultures. We considered the last item necessary because some of the scales measure Hall's (1976) context construct, which by nature must be considered relative to other cultures' communication. In addition, many of the countries are transitional (previously communist), and respondents from this group of countries are much more likely than their parents to have been exposed to globalization in one form or another. Finally, choosing a homogeneous population potentially reduces the individual effects of response styling. As we expected, demographics were homogeneous and representative of traditional college students.

The instrument was carefully translated for both literal and symbolic meaning according to Douglas and Craig's (2006) suggestions. The English version was used only in the United States, the United Kingdom, India, and the Philippines. The survey was administered in each country by a local professor who was asked to obtain a sample of 200 or more students.

Reliability and Unidimensionality

To determine the comparative internal validity of the scales, we examined the following: (1) reliability, (2) measurement model fit, (3) measurement invariance, (4) susceptibility to response styles, and (5) convergent

validity with other measures of culture. As shown in table 4, nearly all the alpha values for the SDM scale are "respectable or better" (i.e., higher than .7), and most of those for the SDA and AL scales are "acceptable" (DeVellis 2003, p. 95). We show that the AL scale is marginal in most cases and that the SDA scale is less than acceptable in Turkey, Mexico, and Finland. Given that reliability is a necessary but insufficient condition for scale reliability, it appears that the SDM scale performs the best. Arguably, the SDA and AL types may perform better with alternative items. However, the initial pretest suggested that even with a relatively large pool of initial items, these scales had marginal reliability across samples, indicating issues with the scale type rather than the specific items. These results are summarized in table 4.

Insert table 4 about here

We analyzed measurement model fit CFA with LISREL 8.8 for each measure both independently and simultaneously. Because the scales attempt to measure the same construct, it is not surprising that significant cross-loadings occurred when we estimated them simultaneously. In general, however, the simultaneous CFAs had good fit measures (without cross-loading specified) for all but Germany (root mean square error of approximation [RMSEA] = .188), India (RMSEA = .09), and Latvia (RMSEA = .107). When estimated independently, the CFA fit was problematic in 13, 7, and 6 countries for the SDA, SDM, and AL measures, respectively (see table 5). Examination of the modification indexes illustrates that in all cases, the low fit is the result of correlated errors. Largely, correlated errors result from wording or meaning redundancy (Netemeyer, Bearden, and Sharma 2003). Although it violates classical test theory, redundancy of items does not inherently invalidate the scale itself. In this case, the violation appears to be minimal.

There are several solutions to the problem of correlated errors. The traditional solution is "parceling" (Bagozzi 1993). This strategy is most useful when the scale contains many items from which to create parcels. Another potential solution is to allow the violation and model the error correlation. Bagozzi (1984) suggests that this solution is justified under certain circumstances, such as when (1) there is a plausible hypothesis for the error correlation, (2) the correlation have negligible impact on the model, (3) the number of correlated errors is kept to a minimum, and (4) they be employed as a last resort. Examination of the issue and data suggests that all these conditions are met. Given the commonality of questions and translations for equivalent meaning, it is not surprising that a few of the items become redundant in some languages. Reexamination of the CFA models that allow a single error to correlate indicates little change in the loadings, and therefore both negligible effect on the model and minimum number of error correlations are met. In all but one case (i.e., SDM scale for Germany), allowing a single error correlation results in the model fit becoming acceptable (we label this $R\chi^2$ in table 5).

Another alternative to the correlated error problem is simply to remove an item from the scale. Correlated errors result from redundancy in wording, meaning, or methods; thus, the removal of the redundant item offers a potential solution. If an item is fully redundant of other items in the scale, thus forcing systematic error to be represented by correlated error terms, removal of the item should not affect the richness of the scale. Different items exhibited redundancy in different countries, so items cannot be removed universally.

Measurement Invariance

Ideally, cross-national/cultural measures should exhibit invariance across samples. This includes configural, metric, and scalar equivalence. Though ideal, full metric invariance rarely is achieved, even in relatively limited two- and three-country groups (Bearden, Money, and Nevins 2006; Mavondo, Babbott, and Tsarenko 2003). As Horn (1991, p. 125) suggests, metric invariance is "a condition to be striven for, not one expected to be fully realized." Horn, McArdle, and Mason (1983) and Steenkamp and Baumgartner (1998) consider metric invariance scientifically unrealistic. This is especially true given larger numbers of groups (He, Merz, and Alden 2008). However, configural and partial metric equivalence is now expected in most cross-national studies.

Configural invariance suggests that the pattern of significant loadings across samples is stable and significant in CFA. Table 4 indicates that with the exception of three single-item instances in the SDA and AL scales and one nonsignificant item in the SDM scale (item numbers in parentheses), all three scales exhibit loading consistency. The three instances of nonconfigural invariance in the SDA scale occurred in conjunction with lower reliability measures. Likewise, two of the three instances in the AL-type scale occurred in countries with marginal reliability. Overall, the patterns suggest reasonable configural invariance, with approximately 3% of the loadings being nonsignificant in the SDA and AL scales and 1% for the SDM scale. In addition, configural variance is similar to the concept of convergent validity (Bagozzi 1981; Kumar, Stern, and Achrol 1992), and significant loadings are considered evidence of convergent validity.

Metric invariance is ideally tested first at the model level by setting all loadings for all samples to be equivalent. However, as a practical matter, a CFA becomes fully saturated with relatively few groups, and thus a multigroup CFA cannot be identified for a 23-country, four- or five-item scale. Thus, we tested metric invariance at smaller group levels (see table 5).

Insert table 5 about here

A brief review of recent scale developments indicates that metric invariance is relatively rare in even small numbers of groups (two to three) (e.g., Bearden, Money, and Nevins 2006). In this case, all three measures exhibited full metric equivalence for at least groups of four countries and as many as six for the SDM scale. Thus, comparatively, all three scales performed well. In all cases, the Japanese and German scales were noninvariant. These are also the countries that would be expected to be at the extremes of the context measure.

Response Style Effects

Both individual and cultural biases in response styles can have a significant impact on cross-cultural scales (Baumgartner and Steenkamp 2001; De Jong et al. 2008). Using Baumgartner and Steenkamp's method (2001), we compared the systematic bias from acquiescence (ARS), disacquiescence (DARS), extreme response style (ERS), response range (RR), midpoint responding (MPR), and noncontingent responding (NCR).

For ARS, DARS, RR, and MPR, we chose ten individual items that had correlations (r) of less than .1 (in most cases, $r < .05$). We used one item from scales that measured uncertainty avoidance, rule adherence, consumers' ethnocentric tendencies (CET), brand loyalty, Internet ability, plagiarism, enjoyment of Internet shopping, importance of teamwork in curriculum, attitude toward the music industry, and ethical views of pirating music from the Internet. For ERS, we choose all these items except rule adherence, CET scale, and Internet ability because the mean of these items significantly deviated from the midpoint by more than 10%.

The NCR style necessitates highly correlated items. We chose six pairs of items with correlations of .636, .670, .729, .746, .750, and .842 from scales measuring Internet ability (two pairs), national identification, cosmopolitanism, CET scale, and brand loyalty. Having tested each country on the biases above, we examined the differential impact of the individual component of response styles on each cultural context scale. That done, we proceeded to isolate and examine the relative effect of individual-level response styles through hierarchical partialing (e.g., Miller et al. 1999) using a series of regression analysis.

Apparently, ARS and MPRS at the individual level did not significantly affect any of the three scale types. In addition, DARS had a negative effect on all scales, with a slightly larger effect on the AL scale than the other two, as we proposed previously. The ERS and RR stylistic responses mostly affected the SDM scale. In this case, it appears that there is a trade-off between NCR affecting both SDA and AL measures of context and ERS and RR affecting the SDM. We believe that NCR is more of a threat to validity than ERS or RR. If desirable, ERS and RR are both easily correctable by increasing the scale from seven to nine points or creating more severe endpoints (e.g., changing from "good" to "great"), while NCR is more difficult to avoid.

NOMOLOGICAL VALIDITY

In this section, we examine the face validity of the measures when compared with previous propositions (including Cardon 2008; Hall and Hall 1990; Koeszegi, Vetschera, and Kersten 2004) on which countries are high and low context. In addition, we examine theoretical/proposed relationships between related constructs to establish and compare convergent validity.

Face Validity

We established partial face validity through the scale development procedure and an examination of the confluence of the item content and the definition of context. In addition, we examined face validity by comparing the outcomes/measures of each scale with expected or previously proposed relative context positioning. As we indicated previously, this has been the Achilles' heel of earlier attempts to measure context.

Insert table 6 about here

Table 6 indicates the relative position on each context scale using factor scores derived from EFA (RelPos). Because of the sample and the use of EFA factor scores, a distribution that differs from what we proposed might be expected because of the slight imbalance of high-context cultures included (about half the sample) relative to middle- and low-context cultures. Thus, relativity must be kept in mind.

The SDM measure provided the best face validity, with all countries falling into a ranking that would be expected. The SDA and AL scales did not perform nearly as well. Of great concern is that the SDA scores for China, Finland, and Mexico were at the far, opposite end of the spectrum of what has been previously proposed. Of less concern was the relatively high position of Slovenia on the SDA measure. Given the similarities of culture between Slovenia and its neighbors, specifically Austria and Italy, we expected them to be relatively low or middle context.

The AL-context measure performed relatively poorly as well. Nine of the countries (almost 40% of the sample) ranked in a manner other than what we expected, though not to the extremes that occurred in the SDA

scale. Only Turkey was the polar opposite of what we expected. A further examination of the response style results and the factor scores indicates that the results from the SDA and AL measures cannot be attributed to differential stylistic responses. In addition, a comparison of the consistency of the measures (table 8) suggests that the Finnish results can be partially explained by simply having poor measures. However, low reliabilities would be more likely to create random noise than to change the relative position of the score.

Convergent Validity

The literature suggests that context should be correlated to individualism/collectivism (COLEC) and perhaps, to a lesser degree, to uncertainty avoidance (UA). Kim, Pan, and Park (1998) suggest that context is related to group orientation, implying a positive relationship between collectivism and context. Other researchers directly propose (Gudykunst and Ting-Toomey 1988) and empirically support this relationship (Kim and Wilson 1994). Likewise, relying on Hall's (1976) original work, Kim, Pan and Park further suggest that low-context people deal better with new or higher risk situations, thus implying a positive association between context and UA. Anecdotally, this correlation might be expected given the high context and high UA of Asian cultures and the opposite in Anglo cultures.

The measures of COLEC and UA from the survey exhibit good reliability, with all but 7 (of 46) measures having alpha values greater than .70 and good fit on the CFA examinations (see reviewer appendix). We used a simple structural equation model from the pooled data to examine to statistical relationships. Pooling of the sample is both ideal and necessary given that cultural measures should be homogeneous within each country and heterogeneous between countries and therefore are necessary to achieve variance within and correlation between measures. In this instance, the use of the structural equation model does not imply causality, but rather is a convenient method for examining convergent validity of latent constructs. Estimates of a pooled structural equation model indicate the structure in figure 1.

Insert figure 1 about here

The results indicate overall good convergent validity of UA with the measures of context. However, the stronger theoretical relationship between COLEC and the context measures suggests some serious problems with the SDA measure as well as a relationship opposite to what we expected with the AL measure (the positive estimate suggests that collectivists are low context because of the manner in which AL is measured—lower context as higher values). The relationship between COLEC and the SDM measure is as we expected.

CONCLUSION

Overall, the new SDM scale performed the best on all criteria. The reliability of the SDM measure was above acceptable in all countries. Although there were slightly more CFA error correlations in the SDM- than SDA-type scale, these are largely correctable issues in implantation by removing an item. The SDA scale was metrically invariant across six countries, which is unusual in the marketing literature. In cases in which researchers cannot correct scales for response styling, the SDM scale also performs better than the others. Finally, and most important, the SDM scale shows the best nomological validity.

Although the importance of context studies is widely accepted, the means to conduct an empirical study of the construct was heretofore limited. With the introduction of the scale developed herein, intercultural

communications researchers will be able to determine discreet differences among countries and better explain their results. Links between context and related topics, such as uncertainty avoidance, collectivism, and power distance (Hofstede 2001), can be scrutinized as well.

An important contribution beyond the scale itself is the introduction of the idea of using a metaphorical semantic differential with nouns instead of the traditional adjectives. This brings with it the advantages of fewer translation errors and more precise conceptual presentation across a variety of countries. Such an approach may pave the way for revisiting previous studies with new instruments that can reconcile conflicting results across various studies. The results here indicate that the new scale type is superior for measuring Hall's (1976) concept of context and might be advantageous for measuring other cultural communication phenomena.

Finally, this research should prove important to practitioners. Previously, marketing professionals needed to deal with only a blurred concept of context. The importance of the subject was undisputed, but the ability to craft individualized approaches based on distinct differences among countries was lacking. As the drive for financial accountability and marketing metrics increases, the need to be able to demonstrate success with marketing initiatives mounts. The new scale should enable communications campaigns, for example, that are tailored to the correct level of context so that maximum returns are achieved. Sales training programs can be modified so that managers can understand the cultures in which they work and can better blend home office demands with foreign communication patterns.

REVIEWER APPENDIX

Construct/Items	Pooled Reliability (Alpha)
Uncertainty Avoidance(UA) **(Adapted from Quintal, Lee and Soutar, 2006)** 1. I avoid taking gambles in life 2. I would rather be safe than sorry 3. I avoid taking chances if possible 4. I like situations that are safe	0.790
Individualist/Collectivist (IndCol) **(Miller et al, 2007)** 1. Group welfare is more important than individual rewards. 2. Individuals should pursue their goals only after considering the welfare of the group 3. I focus on achieving societal goals more than individual accomplishments 4. Group rewards should take priority over individual rewards	0.791

- Quintal, Vanessa, Julie Lee and Geoff Soutar (2006), "Attitudes towards Risk and Uncertainty: Suggested Scales", http://conferences.anzmac.org/ANZMAC2006/documents/Quintal_Vanessa.pdf, accessed July 15, 2009.
- Miller, Chip, Bram Foubert, James Reardon and Irena Vida (2007), "Teenagers' Response to Self- and Other-Directed Anti-Smoking Messages: A Cross-Cultural Study", *International Journal of Market Research*, 49(4), 515-533.

The validity of each of the scales was tested with confirmatory factor analysis (CFA) [Joreskog and Sorbom 1993]. The fit was good (RMSEA 0.56, GFI .98). Convergent validity was tested by examining the t-values of the Lambda-X Matrix (Bagozzi 1981). Ranging from 43.3 to 90.19, all t values were well above the 2.00 level specified by Kumar, Stern and Achrol (1992), indicating high convergent validity. Discriminant validity was examined by setting the individual paths of the Phi Matrix to one and testing the resultant model against the original (Gerbing and Anderson 1988). The high D-squared statistics (Joreskog and Sorbom 1993) implied that the confirmatory factor model fit significantly better than the constrained model for each construct.

REFERENCES

Aaker, J. and Maheswaran, D. (1997). The effect of cultural orientation on persuasion. *Journal of Consumer Research*, *24* (December), 315–28.

Alreck, P. and Settle, R. (1985). *The survey research handbook*. (2nd ed.). Chicago: Richard D. Irwin.

Bagozzi, R. (1981). Evaluating structural equation models with unobservable variables and measurement error: a comment. *Journal of Marketing Research*, *18* (August), 375–81.

——— (1984). A prospectus for theory construction in marketing. *Journal of Marketing*, *48* (Winter), 11–29.

——— (1993). Assessing construct validity in personality research: applications to measures of self-esteem. *Journal of Research in Personality*, *27*, 49-87

Baumgartner, H. and Steenkamp J.B. (2001). Response styles in marketing research: a cross-national investigation. *Journal of Marketing Research*, *38* (May), 143–56.

Bearden, W. O., Money, R.B. and Nevins, J.L. (2006). A measure of long-term orientation: development and validation. *Journal of the Academy of Marketing Science*, *34* (3), 456–67.

Burgess, S. and Steenkamp, J.B. (2006). Marketing renaissance: how research in emerging markets advances marketing science and practice. *International Journal of Research in Marketing*, *23* (4), 337–56.

Cardon, P. (2008). A critique of Hall's contexting model: a meta-analysis of literature on intercultural business and technical communication. *Journal of Business and Technical Communication*, *22* (4), 399–428.

Churchill, G.A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16* (February), 64–73.

Coulter, R., Price, L., Feick, L. and Micu, C. (2005). The evolution of consumer knowledge and sources of information: Hungary in transition. *Journal of the Academy of Marketing Science*, *33* (4), 604–619.

De Jong, M., Steenkamp, J.B., Fox, J.P. and Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: a global investigation. *Journal of Marketing Research*, *45* (February), 104–115.

DeVellis, R. (2003). *Scale development: theory and applications*. (2nd ed.). Thousand Oaks: Sage Publications.

Douglas, S. and Craig, C.S. (2006). On improving the conceptual foundations of international marketing research. *Journal of International Marketing*, *14* (1), 1–22.

_____ (2007). Collaborative and iterative translation: an alternative approach to back translation. *Journal of International Marketing*, *15* (1), 30–43.

Douglas, S. and Njissen, E. (2003). On the use of "borrowed" scales in cross-national research: a cautionary note. *International Marketing Review*, *20* (6), 621–42.

Gerbing, D. and Anderson, J. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, *25* (May), 186–92.

Gidley, J. (2002). Global youth culture: a transdisciplinary perspective. In J. Gidley and S. Inayatullah (Eds.), *Youth futures: comparative research and transformative visions* (pp. 3-18), Santa Barbara: Praeger Publishers.

Gudykunst, W. and Ting-Toomey, S. (1988). Culture and affective communication. *American Behavioral Scientist*, Jan/Feb, 31(3), 384-400.

Hall, E. (1976). *Beyond culture*. New York: Doubleday.

Hall, E. and Hall, M. (1987). *Hidden differences: doing business with the Japanese*. Garden City: Doubleday.

——— and ——— (1990). *Understanding cultural differences: Germans, French, and Americans*. Garden City: Doubleday Press.

He, Y., Merz, M.A. and Alden, D.L. (2008). Diffusion of measurement invariance assessment in cross-national empirical marketing research: perspectives from the literature and a survey of researchers. *Journal of International Marketing*, *16* (2), 64–83.

Hofstede, G. (2001). *Culture's consequences: comparing values, behaviors, institutions and organizations across cultures*. (2nd ed.). Thousand Oaks: Sage Publications.

Horn, J. (1991). Comments on "Issues in Factorial Invariance." In L. Collins and J. Horn, (Eds.), *Best methods for the analysis of change* (pp. 114-125). Washington, DC: American Psychological Association.

Horn, J., McArdle, J. and Mason, R. (1983). When is invariance not invariant? A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist*, *1* (Summer–Fall), 179–88.

Hult, G. T., Ketchen Jr., D., Griffith, D., Finnegan, C., Gonzalez-Padron, T., Harmancioglu, N., Huang, Y., Talay, M.B. and Cavusgil, S.T. (2008). Data equivalence in cross-cultural international business research: assessment and guidelines. *Journal of International Business Studies*, *39* (6), 1027-1044.

International Monetary Fund (2008). World Economic Outlook Database—October 2008. (accessed February 14, 2009), [available at http://www.imf.org].

Johansson, Johny (2003). *Global marketing* (3rd ed.). New York: McGraw-Hill/Irwin.

Joreskog. K. and Sorbom, D. (1993). *LISREL 8: structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software.

Kim, D., Pan, Y. and Park, H.S. (1998). High- versus low-context culture: a comparison of Chinese, Korean and American cultures. *Psychology and Marketing*, *15* (6), 507–521.

Kim, M.S. and Wilson, S.R. (1994). A cross-cultural comparison of implicit theories of requesting. *Communication Monographs, 61*, 47-66.

Kjeldgaard, D. and Askegaard, S. (2006). The globalization of youth culture: the global youth segment as structures of common differences. *Journal of Consumer Research, 33* (2), 231–47.

Koeszegi, S., Vetschera, R. and Kersten, G. (2004). National cultural differences in the use and perception of internet-based NSS: does high or low context matter? *International Negotiation. 9* (1), 79–109.

Kroll, J. and de Groot, A. (1997). Lexical and conceptual memory in the bilingual: mapping form to meaning in two languages. In A. de Groot and J. Kroll, (Eds.), *Tutorials in bilingualism: psycholinguistic perspectives* (pp. 169-199). Mahwah: Lawrence Erlbaum Associates.

Kumar, N., Stern, L., and Achrol, R. (1992). Assessing reseller performance from the perspective of the supplier. *Journal of Marketing Research*, *29* (May), 238–53.

Luna, D., Ringberg, T. and Peracchio, L. (2008). One individual, two identities: frame switching among biculturals. *Journal of Consumer Research*, *35* (August), 279–93.

Mavondo, F., .Babbott, M. and Tsarenko, Y. (2003). Measurement invariance of marketing instruments: an implication across countries. *Journal of Marketing Management*, *19* (June), 523–40.

Miller, C., .Foubert, B., Reardon, J. and Vida, I. (2007). Teenagers' response to self- and other-directed anti-smoking messages: a cross-cultural study. *International Journal of Market Research*, *49* (4), 515–33.

Miller, C., .Reardon, J. and McCorkle, D. (1999). The effects of competition on retail structure: an examination of intratype, intertype, and intercategory competition. *Journal of Marketing*, *63* (October, 4), 107-120.

Monga, A.B. and John, D.R. (2007). Cultural differences in brand extension evaluation: the influence of analytic versus holistic thinking. *Journal of Consumer Research*, *33* (March), 529–36.

Netemeyer, R.G., Bearden, W.O. and Sharma, S. (2003). *Scaling procedures: issues and applications*, Thousand Oaks: Sage Publications.

Nguyen, A., Heeler, R.M. and Taran, Z. (2007). High-low context cultures and price-ending practices. *Journal of Product & Brand Management*, *16* (3), 206–214.

O'Donovan, D. (1965). Rating extremity: pathology or meaningfulness? *Psychological Review*, *72* (5), 358–72.

Ohashi, R. (2000). High/low context communication: conceptualization and scale development. *Dissertation Abstracts International*, *61* (5), 1680, University Microfilms No. 9971973.

Richardson, R. and Smith, S. (2007). The influence of high/low context culture and power distance on choice of communication media: students' media choice to communicate with professors in Japan and America. *International Journal of Intercultural Relations*, *31* (4), 479–501.

Steenkamp, J.B. and Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25* (1), 78–90.

Strizhakova, Y., Coulter, R. and Price, L. (2008). The meanings of branded products: a cross-national scale development and meaning assessment. *International Journal of Research in Marketing*, *25* (2), 82–93.

Trompenaars, F. (1994). *Riding the waves of culture: understanding diversity in global business*. New York: Richard D. Irwin.

Table 1
Preliminary Scales for Context Measurement

Scale Type	Scale Items	Acronym
Semantic Differential Adjective (SDA)	**Communicating in my native language is:**	
	Implicit/Explicit	**SDA1Explic**
	Not Detailed/Detailed	**SDA2Detail**
	Indirect/Direct	**SDA3Direct**
	Abstract/Defined	**SDA4Defined**
	Assumptive/Complete	**SED5Comple**
SDM	**Communicating in my native language is most like:**	
	Precision Watch/Rhythms of Nature	**SDM1Nature**
	Mathematics/Poetry	**SDM2Poet**
	Physics/Art	**SDM3Art**
	Concrete/Flowers	**SDM4Flowe**
	Engineering/Ballet	**SDM5Ballet**
	Laser Beam/Sunlight	**SDM6Sun**
Anchored Likert (AL)	**In general, {NATIONALITY} …**	
	tend to communicate in very explicit language.	**AL1Explic**
	say what we mean and mean what we say.	**AL2Mean**
	leave little room for interpretation of what we say.	**AL3Interp**
	express our ideas in very precise terms.	**AL4Precise**
	words should be interpreted dependent on the situation.	**AL5Depend**

Table 2
Pretest Comparison

Country	**N**	**SDA**	**SDM**	**AL**
France	89	$\alpha_a = .77$; $\alpha_b = .77$	$\alpha_a = .78$; $\alpha_b = .78$	$\alpha_a = .36$; $\alpha_b = .60$
Italy	128	$\alpha_a = .73$; $\alpha_b = .72$	$\alpha_a = .81$; $\alpha_b = .85$	$\alpha_a = .56$; $\alpha_b = .73$
United States	152	$\alpha_a = .75$; $\alpha_b = .79$	$\alpha_a = .83$; $\alpha_b = .82$	$\alpha_a = .59$; $\alpha_b = .69$
Russia	104	$\alpha_a = .74$; $\alpha_b = .76$	$\alpha_a = .81$; $\alpha_b = .84$	$\alpha_a = .64$; $\alpha_b = .79$
Philippines	92	$\alpha_a = .82$; $\alpha_b = .77$	$\alpha_a = .84$; $\alpha_b = .83$	$\alpha_a = .82$; $\alpha_b = .80$

Notes: α_a = initial alpha, and α_b = alpha after purification.

Table 3
Country/Sample Description

Country	**Hall's Context**	***Language Root***	**PPP per Capita**	**Hofstede's Dimensions**		
				Individualism	Masculinity	UA
Belgium		*Germanic*	35,388	75	54	94
Croatia		*Slavic*	16,754			
Finland	Lc*Low	*Finno-Ugric*	35,349	63	26	59
France	Middle	*Romance*	33,509	71	43	86
Guatemala	High	*Romance*		6	37	101
Germany	Lc*Lowest	*Germanic*	34,212	67	66	65
Italy	Middle	*Romance*	30,365	76	70	75
Latvia		*Baltic*	17,488			
Lithuania		*Baltic*	17,733			
Portugal		*Romance*	21,779	27	31	104
Russia	High*	*Slavic*	14,705	39	36	95
Serbia		*Slavic*	10,071			
Slovenia		*Slavic*	27,227			
United Kingdom	Middle	*Germanic*	35,634	89	66	35
United States	Low	*Germanic*	45,725	91	62	46
Mexico	High	*Romance*	14,120	30	69	82
China	Highest	*Sinitic*	5,325	20	66	30
India	High*	*Indo-Iranian*	2,563	48	56	40
Japan	Highest	*Japanese*	33,596	46	95	92
Kazakhstan		*Slavic/Turkic*	10,837			
Philippines		*Indonesian*	3,383	32	64	44
Tunisia	High	*Arabic*	7,535			
Turkey		*Turkic*	12,858	37	45	85
RANGE		*10 groups*	2.6-45.7K	6-91	26-95	30-104

Notes: PPP per capita figures from International Monetary Fund (2008). UA = uncertainty avoidance. Adapted from Koeszegi, Vetschera, and Kersten (2004).

Table 4
Initial Results

Country	N	SDA			SDM			AL		
		α	CFA Fit χ^2 **($R\chi^2$)**	Low t-value	α	CFA Fit χ^2 **($R\chi^2$)**	Low t-value	A	CFA Fit χ^2 **($R\chi^2$)**	Low t-value
Belgium	249	.800	5.98	7.98	.879	4.23	11.74	.771	**7.83* (.14)**	7.80
China	207	.836	**6.11*** (.19)	10.87	.780	**16.73* (.36)**	4.25	**.688**	.46	4.09
Croatia	205	.730	**7.24* (2.91)**	7.20	.863	3.10	8.53	**.632**	4.36	5.71
Finland	223	**.413**	**10.64* (1.39)**	3.15	.821	2.53	7.53	.729	3.69	7.09
France	329	.725	2.13	7.17	.783	**13.04* (.61)**	9.91	**.595**	.59	5.29
Guatemala	238	.743	**13.46* (2.02)**	6.58	.871	2.28	11.42	**.588**	4.15	5.68
Germany	205	.885	**13.45* (.38)**	21.03	.962	**186.7* (19.2*)**	23.11	**.691**	.45	**.25(3)**
India	193	.836	1.20	5.46	.806	4.44	4.89	**.682**	.94	12.06
Italy	409	**.654**	**21.18* (2.72)**	8.46	.847	**18.96* (3.57)**	15.49	.734	4.70	8.68
Japan	257	.779	**18.70* (3.96)**	8.63	.776	4.71	7.79	**.674**	3.85	6.31
Kazakhstan	372	**.691**	.95	**.02(4)**	.837	1.79	2.41	.792	**7.04* (1.76)**	**1.09(4)**
Latvia	123	.765	**6.08* (.12)**	5.53	.743	1.62	6.07	**.620**	.76	3.51
Lithuania	196	.741	**9.96* (1.61)**	6.47	.848	.33	10.40	**.674**	**6.18* (.91)**	4.7
Mexico	215	**.584**	1.33	**1.63(1)**	.812	.27	7.38	.733	**9.36* (2.24)**	7.45
Philippines	379	.868	**11.72* (.48)**	16.62	.833	**13.92* (1.65)**	13.41	.795	1.43	12.69
Portugal	291	.875	2.49	14.51	.842	.90	10.45	.729	.025	8.06
Russia	332	.750	**8.42* (.34)**	7.64	.844	3.76	11.16	**.609**	**7.39* (2.30)**	5.42
Serbia	253	.752	1.88	8.14	.799	**16.09* (.94)**	7.17	**.652**	1.05	6.54
Slovenia	291	.730	2.25	9.08	.837	2.27	10.24	**.529**	.05	2.30
Tunisia	231	**.624**	3.80	3.88	.794	2.18	8.29	**.604**	2.00	5.32
Turkey	337	**.468**	4.18	**1.37(2)**	**.684**	**7.63***	6.44	**.609**	3.44	5.30

						(2.85)				
United Kingdom	203	.869	**14.39* (.82)**	11.66	.839	5.71	9.81	.794	**9.16* (.17)**	9.05
United States	448	.854	**9.55* (2.83)**	13.96	.858	.86	11.42	.735	3.89	10.33
OVERALL/POOLED		.779			.867			.731		

Notes: # ns loads = number of insignificant factor loadings. **($R\chi^2$) = revised model chi-square.**

Table 5
Groups Exhibiting Full Metric Equivalence

	SDA Scale	SDM Scale	AL scale
Groups displaying full metric measure invariance at $p < .05$ unless noted	BEL, FRA, ITA, TUN $\chi^2 = 14.25$ d.f. = 12	US, LT, POR, FRA, ITA, SLO $\chi^2 = 28.61$ d.f. = 25	BEL, LIT, PHI, ITA $\chi^2 = 11.42$ d.f. = 12
	CRO, LAT, LIT, IND $\chi^2 = 12.24$ d.f. = 12	RUS, MEX, UK, PHI $\chi^2 = 17.16$ d.f. = 15	UK, SERB, POR, FIN $\chi^2 = 7.48$ d.f. = 12
	SLO, SERB, UK $\chi^2 = 7.45$ d.f. = 8	FIN, SER, LAT $\chi^2 = 11.54$ d.f. = 10	UK, SERB, KAZ, MEX $\chi^2 = 17.6$ d.f. = 12
	POR, USA $\chi^2 = 8.19$ d.f. = 4	{IND, PRC} TUR $\chi^2 = 4.18$ d.f. =8 $\chi^2 = 19.16$ d.f. = 12 w/ Turkey ($p < .10$ with Turkey)	FRA, JAP $\chi^2 = 8.91$ d.f. = 4
	KAZ, PRC $\chi^2 = 5.36$ d.f. = 4	BEL, KAZ $\chi^2 = 11.80$ d.f. = 5 ($p < .10$)	TUN, IND $\chi^2 = 6.72$ d.f. = 4
		TUN, CRO $\chi^2 = 8.15$ d.f. = 5	TUR, RUS $\chi^2 = 10.84$ d.f. = 4 $p < .10$
Noninvariant groups	PHI, MEX, RUS, GER, TUR, FIN, JAP	JAP, GER	PRC, GER, SLO, LAT, USA, CRO

Notes: χ^2 = model difference chi–square.

Table 5
Face Validity: Relative Scale Position

Country	Expected Position[1]	SDA		SDM		AL	
		Relative Position	Factor Score	Relative Position	Factor Score	Relative Position	Factor Score
Japan	Very High	Highest	–.704	Highest	.428	Highest	–.770
China	Very High	Low	.622	High	.317	High	–.234
Guatemala	High	Low	.127	High	.218	Mid	–.076
India	High	High	–.154	High	.287	Mid	.055
Mexico	High	Low	.206	High	.239	High	–.329
Russia	High	VHigh	–.473	High	.281	High	–.403
Tunisia	High	Mid	–.093	High	.129	High	–.178
Turkey	High	High	–.294	High	.226	Low	.269
France	Middle	Mid	.091	High	.384	High	–.142
Italy	Middle	Mid	–.098	High	.416	Mid	–.048
United Kingdom	Middle	Low	.236	High	.134	Low	.240
Finland	Low	High	–.321	Low	–.377	Low	.248
United States	Low	Mid	.061	Low	–.631	Mid	–.030
Germany	Lowest	Lowest	1.276	Lowest	–2.06	Lowest	1.30
Portugal	*High*	Low	.501	High	.260	Low	.220
Belgium	*Low*	Low	.226	Low	–.117	High	–.256
Croatia	*Low*	Low	.190	Low	–.774	Lowest	1.28
Philippines	*High*	Mid	–.053	Mid	.096	Low	.309
Latvia	*Low*	Low	.274	Low	–.167	High	–.247
Lithuania	*Low*	Low	.209	Mid	–.060	High	–.166
Serbia	*High*	High	–.186	Mid	.013	High	–.313
Kazakhstan	*High*	Very High	–.477	High	.271	High	–.375
Slovenia	*Low*	High	–.239	Low	–.223	Mid	.043
Number Unexpected		4/*2*		0		4/*5*	

Notes: We derive expected positions from Hall and Hall (1990), Cardon (2008), and Koeszegi, Vetschera, and Kersten (2004) when possible. When not specified by previous research, we and colleagues proposed positions based on anecdotal evidence and opinions of colleagues familiar with those countries (in Italics). We expected Belgium to be low in context because the measure was taken in the Flemish (Antwerp) part of the country.

Figure 1

CONVERGENT VALIDITY – T-TESTS