Descriptive Analytics: It's Power to Test the Applicability of Cross-National Scales in Exploratory Studies

Srinivas Durvasula
Professor and Edward A. Brennan Chair in Marketing
Marquette University
Milwaukee, WI 53233
Mail: Srinivas.durvasula@marquette.edu
USA

Steven Lysonski
Professor and Miles Research Scholar
Marquette University
Milwaukee, WI 53233
USA

Abstract:

Conventional methodology for validating measures in consumer research relies on structural equation modeling. But, this procedure requires a fairly large sample size and a clear conceptualization of the relationship between individual items and various scale dimensions. Neither of these requirements may be met in exploratory cross-national studies. Hence, this paper addresses scale validation issues in exploratory cross-national research where sample size is a major concern. Specifically, it uses cross-national data on the vanity measure as an exemplar and a battery of descriptive analytics to show how to assess scaling assumptions, reliability, and dimensionality of consumer behavior measures. The scale validation procedure we describe in this paper has implications for researchers who use multi-item rating scales as measures of consumer behavior constructs.

Keywords: Cross-cultural, Scale Validation, Exploratory Research, Cross-national, Scale Applicability

1. Introduction:

International business studies typically involve application of consumer behavior measures to investigate cross-national differences (Halkias, Davvetas, and Diamantopoulos 2016; Durvasula and Lysonski 2015). While some of those studies may have used established measures (e.g., CETSCALE, SERVQUAL), others may be exploratory in nature. In either case, researchers use multi-item rating scales based on the likert scale format to measure the underlying construct. Consumer responses are then summed (or averaged) to form an overall score. The validity of measurement scales, however, is based on certain assumptions (e.g., internal consistency, external consistency, and uni-dimensionality). When those scaling assumptions are not met, we have no way of knowing whether observed cross-national consumer differences on the summed scores are due either to translation problems, country-specific differences in the definition of the consumer behavior construct, or due to true consumer differences on the underlying construct. Hence, it is crucial to design valid cross-national measures that satisfy underlying scaling assumptions (Steenkamp and Baumgartner 1998).

Conventional methodology for assessing cross-national scale applicability calls for an application of confirmatory factor analysis and the use of structural equation modeling (SEM) (Steenkamp and Baumgartner 1998). However, it has certain limitations; the most notable being that SEM is sample-intensive. For cross-national researchers, such limitations may be likely in that scale analysis commonly takes place first in pilot or exploratory studies where sample sizes can be rather small (Netemeyer, Durvasula, and Lichtenstein 1991). Some of the well-documented limitations of SEM are non-convergence

and improper solutions (Anderson and Gerbing 1985; Boomsma 1985), as well as bias in the estimated factor loadings and standard errors (Anderson and Gerbing 1985). Evidence suggests that this bias in parameter estimates exists in small samples; irrespective of the estimation procedure (e.g., maximum likelihood) used (Benson and Felishman 1994; Dolan 1994). Furthermore, in exploratory cross-national studies, researchers may be performing scale analyses with not so well-defined expectations about the structure of the items and their relationships to various measures. A key objective of those studies may be to determine whether a set of items forms a uni-dimensional scale consistently across various countries. As such, there is a need for a method that is easier to use vs. SEM in exploratory research or in studies that are based on small samples.

Often, researchers who work with small samples, but who are unfamiliar with SEM, often rely on reliability analysis and exploratory factor analysis.Yet, reliability analysis is restricted to assessing internal consistency because only items from a single scale are considered. Internal consistency addresses only one of the measurement issues. The other major issue is dimensionality, which is not addressed by reliability analysis. In contrast, exploratory factor analysis is designed to assess items from multiple scales. But, as the technique is not rooted in a measurement model such as the classical test model (Saris and Hartman 1990), it only provides an indirect test of uni-dimensionality. Moreover, as exploratory factor analysis does not permit researchers to constrain items to load on specific factors, results based on exploratory factor analysis can even be misleading (Steenbergen 2000).

In sum, existing approaches for assessing scaling assumptions in exploratory cross-national research studies are inadequate, particularly when small sample sizes are a concern. The goal of this paper is to develop a tool kit of descriptive analytics that help assess scaling assumptions and the cross-national applicability of measures used in consumer research. We believe that our paper adds to the research stream in cross-national studies that focuses on measurement issues. In the remainder of the paper we explain the importance of dimensionality in cross-national research, outline the descriptive analytics based approach for assessing cross-national scale applicability, present the results based on an analysis of four-country data on vanity, and conclude with a discussion of the proposed approach.

2. The Importance of Measure Dimensionality in Cross-National Research:

Dimensionality of consumer behavior measures is dependent on the behavior of individual scale items. Scale items are uni-dimensional if they satisfy two conditions-*internal consistency* and *external consistency*. For a scale to be internally consistent, scale items should be associated with each other. Further, the correlations among scale items must be attributable entirely to their association with a common underlying construct or dimension (Gerbing and Anderson 1988). External consistency, on the other hand, implies that no scale item should tap more than one construct. Any correlation between items from different scales then can be attributed entirely to the correlation between the underlying constructs (Gerbing and Anderson 1988). To establish scale dimensionality, it is therefore imperative to examine both internal consistency and external consistency of scale items.

Establishing dimensionality of consumer behavior measures, in turn, is of paramount importance in cross-national research (Clark and Watson 1995; Netemeyer, Bearden, and Sharma 2003). In the process of operationalizing latent constructs, researchers often use composite scores by summing or averaging across items designed to measure the construct of interest. The application of such scores is only meaningful if the items have uni-dimensionality. When multidimensional scales are treated as uni-dimensional (i.e., summed or averaged item composites), they could result in interpretational ambiguities. In other words, if a construct were to be multidimensional, but all item scores were to be summed/averaged across dimensions into a single composite score and correlated with a criterion variable, such a correlation would at best ambiguous and at worst, misleading (Durvasula et al., 2006).

Neuberg, West, Thompson, and Judice (1997) presented a persuasive case as to why multidimensional

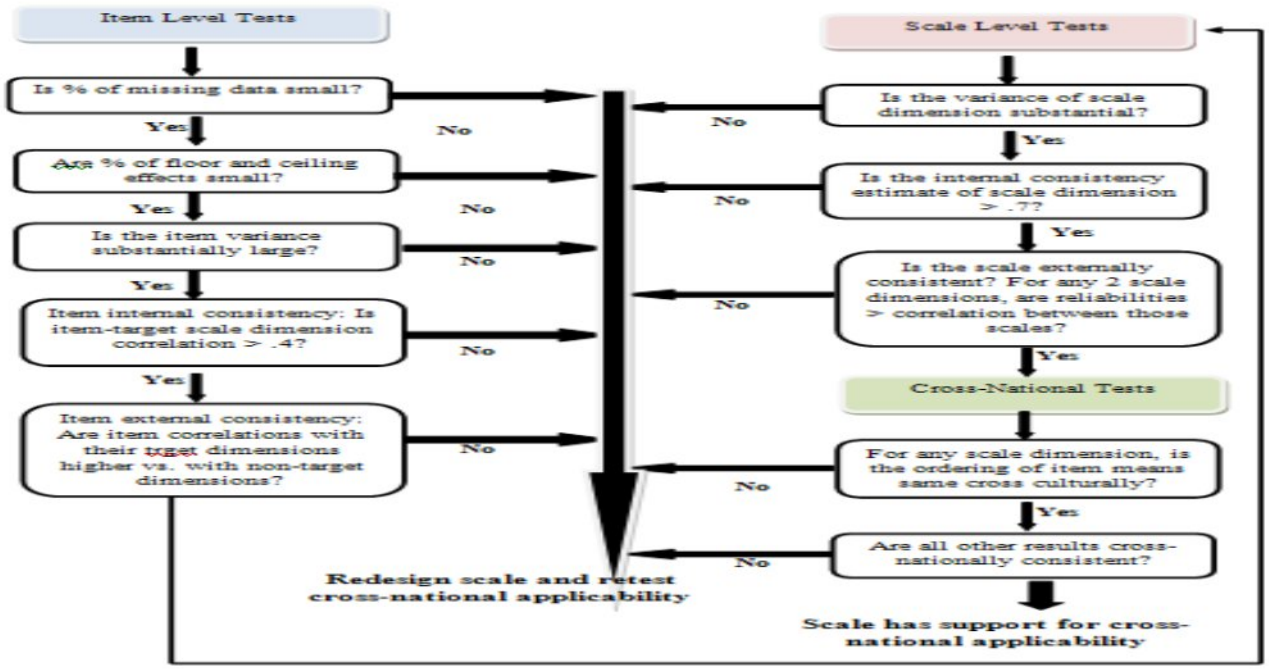
scales should not be treated as if it they are uni-dimensional. Of critical importance to cross-national research, if the dimensionality of a scale varies from one country to the other, any mean comparisons based on composite scores would produce worthless results. As such, establishing dimensionality of measures is a necessary condition for internal consistency, construct validity, and model testing. When measures exhibit validity in various countries, then they become cross-nationally applicable. Hence, applicability presupposes validity. The preferred method for establishing cross-national applicability of measures, and the only one that most scholars in consumer research are acquainted with, is SEM. The following section provides an alternative method for establishing scale validity, one that is more appropriate if researchers are confronted with small samples, especially in exploratory studies.

3. An Alternative Approach for Establishing Scale Applicability:

While SEM offers a strong test for measure applicability, its application in small sample cross-national research studies may be limited because of non-convergence of parameter estimates, empirical under-identification of factor models, and improper solutions. In contrast, exploratory factor analysis may seem like a more appropriate method that is tailored to small-sample studies, but this method does not permit researchers to specify or constrain as to which items should be associated with what scale dimension. Therefore, the factor in a factor analysis represents a statistical construct, but it cannot be thought of as a psychological construct. For assessing scale dimensionality, however, the ability to specify which items relate to what constructs is important. What then is a feasible alternative? The answer can be found in the works of Likert.

Likert (1932) listed key assumptions of summated rating scales ones that must be met before a scale can be applied to examine group differences. Using a battery of descriptive analytics Ware and Gandek (1998) showed how to test those likert scale assumptions for measures used in life sciences. We adapt that method and propose it as an alternative approach to SEM for establishing cross-national applicability of consumer behavior measures especially where sample sizes are relatively small. The procedure involves examining a measure at the item level as well as at the scale level. If the results are supportive, and if they are consistent across countries, then the underlying measure will have cross national applicability. The outline of this alternative approach to SEM is shown in Figure 1 and further illustrated below.

Figure 1: Steps for Establishing Cross-National Scale Applicability in Small Studies



3.1 Assess Item Quality using Item-Level Descriptive Analytics:

3.1.1 Missing and Out-of-range Values. If a large amount of data is missing, then it is impossible to measure the underlying concept with confidence. Instead, it indicates the possibility that the respondents did not understand how to respond to the scale items, or the likelihood that they had difficulty with wording of the scale items.

3.1.2 Floor and Ceiling Effects. These effects determine whether all of the response choices of a scale are used, or if consumers deliberately chose certain response styles such as consistently checking only the scale end points. When consumers exhibit different response styles in different countries, then it is impossible to determine whether or not they actually differ on the underlying concept, making any further cross-national comparisons futile. While the floor effect indicates dis-acquiescence response style – the tendency of respondents to strongly disagree with a statement (i.e., choose a response of 1 or 2 of a 7-point scale item), the ceiling effect indicates the acquiescence response style, that is, it shows what percent of respondents strongly agreed with a statement (i.e., selected the options of 6 or 7 on a 7-point scale). If extreme response styles are more prevalent in some countries than the others, then the researcher is confronted with response style bias in those countries. It then becomes impossible to determine real differences across countries on the underlying construct.

3.1.3 Item Means. As described by Ware and Gandek (1998), under traditional likert scaling criteria, means of individual scale items should be fairly equal. This will usually be the case if all scale items are tapping the same part of the construct domain. However, if the researcher uses different items to tap different aspects of the concept domain, then it is possible and acceptable to have non-equivalent item means for some of the scale items. More importantly, while the average score of individual scale items is expected to vary with the level of the underlying construct (e.g, high, medium, or low level of vanity) and with the populations sampled, the placement order of item means and the approximate differences between them should not vary across countries. Otherwise, if placement order of item means of a scale were to vary significantly across countries, then it calls into question the cross-national applicability of the corresponding measure.

3.1.4 Item Variance. The best items of a measure are those that exhibit significant variability, as they play a useful role in detecting differences across consumer groups in terms of the underlying concept. Also, variances (or standard deviations) of individual items should be roughly equal. However, it is not always possible to obtain high item variances. In such cases, it may still be appropriate to include an item as part of a measurement scale, if the reason for its inclusion is to tap an important part of the construct domain that is not being measured by other scale items.

3.2 Multi-Trait/Multi-Item Correlation Matrix:

The multi-trait item-scale correlation matrix provides information to test a number of scale assumptions that affect cross-national applicability of a measure – internal consistency, equality of item-scale correlation, and external consistency.

3.2.1 Item Internal Consistency. It measures the correlation of each scale item with the composite or sum score of all the remaining items in the scale. Item internal consistency can be deemed satisfactory if an item correlates 0.40 or more with its hypothesized scale. However, for items whose correlation with the sum score is below 0.40, whether or not to delete those items from the scale depends on how critical those items are to capturing the true domain of the underlying concept.

3.2.2 Equality of Item-Scale Correlations. Items in a scale should have fairly similar correlations with the composite or sum score of the scale. Otherwise, a low correlation implies that the corresponding scale item does not contribute equal proportion of information to the associated measure. Items that do not

contribute enough information should then be excluded. The remaining items should all be given the same weight when computing the composite or sum score. There is strong empirical support for this practice so long as items have approximately equal correlations with their target scales (Armor 1974; Ware and Gandek 1998). However, when all items contribute significantly to the total score, this standard or equality of item-scale correlations can be considered satisfied, even if item-scale correlations vary (from 0.40 to 0.70 or more).

3.2.3 Item External Consistency (i.e., Item Discriminant Validity). It is not enough to show that an item measures the concept it is supposed to measure. For discriminant validity, it is also important to show that the item does not correlate highly with measures of other concepts. The multi-item multi-trait correlation matrix can be used to compare the correlation of an item with its hypothesized scale to the correlation of the same item with all the other scales. For external consistency or discriminant validity, the correlation of a scale item with composite scores of other scales should be smaller (vs. correlation of the scale item with other items of the same concept).

3.2.3 Item Similarities. An even more systematic way of assessing scale dimensionality is by examining item similarities. Similarity coefficients are derived from the statistical consequences of uni-dimensionality, so they provide a better way of evaluating scale dimensionality as compared to exploratory factor analysis (Steenbergen 2000). The greater the similarity between two items as measured by the similarity coefficients, the better those items fit the classical test model, and the more valid the conclusion is that they form a uni-dimensional scale.

3.3 Assessing Overall Scale Quality Based on Scale-Level Analytics:

3.3.1 Scale-Level Descriptive Analytics. After item-level analytics in each country, the next step is to apply scale level analytics. This involves comparing scale means, standard deviations, and floor and ceiling values. In the event significant differences are found in mean scale scores across cross-nationally comparable samples, one should perform further analysis to determine if the differences are due to translation problems or to country-specific differences on the underlying construct. Similar to the expectation that individual items should have high variance, scale scores should also have high variability. As explained by Ware and Gandek (1998), this requirement is even more crucial for scale validity.

3.3.2 Scale Internal Consistency. The average of all inter-item correlations within a scale points to the internal consistency of a measure. A minimum reliability level of 0.70 has been suggested (Nunnally and Bernstein, 1994) for acceptable scale reliability.

3.3.3 Correlations between Scales. To evaluate how distinct each scale is from other scales, correlations among all scales are computed and compared with reliability estimates (Campbell and Fiske, 1959), where the reliability coefficient can be viewed as a correlation between a scale and itself. To the extent that the correlation between two different scales is less than their respective reliability coefficients, there is evidence that each of the two scales possess discriminant validity. In contrast, when the correlation between two scales is close to their respective scale reliabilities, then those scales lack discriminant validity. Instead, they can only be viewed as alternate form measures of the same concept. The scale reliabilities and inter-scale correlation thus help us determine whether or not the two scales have discriminant validity.

In sum, if the battery of descriptive analytics provides consistent results cross-nationally as described above, then there is support for cross-national applicability of the underlying measure. Individual items of the scale can then be summed and the composite score to examine cross-national mean differences. In the next section we describe the vanity measure and our cross-national data set to illustrate how to assess cross-national validity of a measure.

Method:

The vanity measure is a scale consisting of 21 scale items. They collectively assess four distinct yet related concepts of vanity (Netemeyer, Burton, and Lichtenstein, 1995). The various scale items of this measure were obtained using the likert scale. Figure 2 shows a description of the vanity scale. This scale was applied in four countries, the United States, New Zealand, China, and India. While the U.S. and New Zealand are developed countries representing the Western cultures, China and India are developing countries that represent the Eastern cultures. An average of 100 respondents completed the vanity scale across the four countries. While the survey in China was administered in Chinese, the English version was administered in the other three countries. Appropriate translation procedure was employed to convert the original English version of the survey into Chinese. Across all four countries, young adults with similar educational background completed the survey. As opposed to random samples, comparable samples such as young adult samples are necessary to facilitate cross-national comparisons.

Figure 2
Items Measuring Vanity

Physical-Concern Items

The way I look is extremely important to me (PC1)
I am very concerned about my appearance (PC2)
I would feel embarrassed if I was around people and did not look my best (PC3)
Looking my best is worth the effort (PC4)
It is important that I always look good (PC5)

Physical-View Items

People notice how attractive I am (PV1)
My looks are very appealing to others (PV2)
People are envious of my good looks (PV3)
I am a very good-looking individual (PV4)
My body is sexually appealing (PV5)
I have the type of body that people want to look at (PV6)

Achievement-Concern Items

Professional achievements are an obsession with me (AC1)
I want others to look up to me because of my accomplishments (AC2)
I am more concerned with professional success than most people I know (AC3)
Achieving greater success than my peers is important to me (AC4)
I want my achievements to be recognized by others (AC5)

Achievement-View Items

In a professional sense, I am a very successful person (AV1)
My achievements are highly regarded by others (AV2)
I am an accomplished person (AV3)
I am a good example of professional success (AV4)
Others wish they were as successful as me (AV5)

4. Empirical Illustration:

For illustration purposes, we followed the descriptive analytics procedure as described in Figure 1 to assess cross-national applicability of the vanity scale. Following is a summary of the results.

4.1 Item-Level Descriptive Statistics:

4.1.1 Missing Values. As shown in Table 1, across the four samples, the percentage of missing values is generally very small for all vanity scale items. Only in the case of 3 items out of 21 (AC1, AC2, and AV4), that too limited to the Indian sample, the missing value percentage exceeded 5%.

4.1.2 Item Variance. In general, items exhibited greater variability in India and China as compared to the U.S. and New Zealand. For an item measured on a 5-point rating scale, it is desirable to have a standard deviation of about 1. So, for a 7-point rating scale, this value should be above 1. Compared to this recommended value, the standard deviation is somewhat low for 3 (out of 21) items in the U.S. sample. However, for 13 of the items in the U.S. sample, the standard deviation is well in excess of 1.

4.1.3 Mean Values. When making cross-national comparisons we should look for similarity of item means and whether those values are ordered in a roughly similar fashion across the samples. An inspection of item means for the physical view dimension, PV1 has the highest mean and PV3 has a fairly low mean value across the four samples. As for the physical concern and achievement concern dimensions, all item means are above 4. The only exception is the mean score for PC3 in the U.S. sample. Another similarity across the samples is that among items representing achievement view, AV4 and AV5 have the smallest mean. In sum, there appears to be a semblance of order among item means across the four samples.

4.1.4 Item Floor and Ceiling Values. The term "floor" represents selection of the lowest response category (e.g., strongly disagree), whereas the term "ceiling" represents selection of the highest response category (e.g., strongly disagree). High ceiling values suggest the possibility of acquiescence response bias. A high ceiling value coupled with a high floor value suggests the possibility of extreme response style bias. For 7-point rating scales, when there is a symmetric distribution of responses, then we would expect 14% of respondents to select each response category. So, floor or ceiling values far in excess of 14% and a combined floor and ceiling value above 28% raise concern about significant response patterns. Further, sizeable differences in those values across samples imply that scale responses are affected by response style biases.

Results of floor and ceiling percentages are provided in Table 2. It is evident that in India and China, there is a greater likelihood of strongly agreeing with physical concern and achievement concern items. That is why the ceiling percentages are fairly high. However, it can be argued that globalization and the impact of global media have significantly raised concern for physical appearance in India and China. The intense job competition in these two countries, among other factors, is likely to have raised concern for professional achievements in India and China. Given such possibility, perhaps high ceiling percentages for physical concern and achievement concern related items is not unusual in India and China. Moreover, the mean responses to items representing physical concern and achievement concern aren't significantly higher in these two countries as compared to the U.S. and New Zealand, where extreme responses are not as common. It is for the same reason that the low floor and ceiling percentages in the U.S. and New Zealand, which otherwise would have suggested middle response bias, also present no major measurement issues.

Table 1
Item Level Descriptive Statistics

	New Zealand			India			China			United States		
	Mean	Std Dev	%miss	Mean	Std Dev	%miss	Mean	Std Dev	% miss	Mean	Std Dev	% miss
Phy Concern												
PC1	5.46	1.1	0	5.88	1.33	0	5.33	1.37	0	5.44	0.97	0
PC2	5.16	1.13	0	5.49	1.44	0	4.79	1.38	0	5.32	1.07	0
PC3	4.25	1.38	0.57	5.13	1.51	0.85	4.69	1.46	1.64	3.83	1.42	0
PC4	5.13	1.13	0.57	5.04	1.36	3.39	5.24	1.45	0.82	5.05	0.97	1.18
PC5	4.54	1.27	1.14	5.05	1.55	0	5.01	1.4	0	4.2	1.24	0
Phy View												
PV1	4.14	1.27	0.57	4.51	1.6	2.54	4.09	1.38	2.46	4.25	1.02	0
PV2	4.01	1.15	0	4.48	1.6	2.54	3.89	1.47	1.64	4.12	0.94	1.18
PV3	3.57	1.22	1.14	3.57	1.8	3.39	3.61	1.42	3.28	3.49	1.28	0
PV4	3.88	1.26	1.71	4.41	1.51	1.69	3.71	1.56	0.82	4.19	1.19	0
PV5	3.81	1.34	1.14	3.54	1.75	2.54	3.41	1.57	1.64	4.1	1.27	0
PV6	3.69	1.35	1.71	3.56	1.69	1.69	3.72	1.54	2.46	3.9	1.27	0
Achiev Concern												
AC1	4.66	1.29	1.71	4.91	1.89	5.08	5.04	1.42	3.28	4.63	1.32	0
AC2	4.96	1.32	1.71	4.85	1.87	5.93	4.92	1.71	1.64	5.1	1.18	0
AC3	4.72	1.34	1.14	5.63	1.61	2.54	4.74	1.65	0	4.35	1.54	2.35
AC4	4.67	1.3	2.29	5.61	1.59	1.69	5.44	1.38	1.64	4.41	1.36	0
AC5	5.07	1.22	2.29	5.39	1.68	0.85	5.13	1.56	0	5.31	0.94	0
Achiev View												
AV1	4.37	1.06	1.71	4.62	1.46	3.39	3.86	1.48	0	4.84	1	1.18
AV2	4.28	1.06	1.14	4.57	1.61	2.54	3.86	1.43	1.64	4.59	0.99	0
AV3	4.49	1.14	2.29	4.46	1.72	3.39	3.8	1.54	1.64	4.96	1	0
AV4	3.87	1.14	1.71	3.66	1.68	5.93	3.47	1.59	2.46	4.31	1.11	0
AV5	3.81	1.21	2.86	3.93	1.79	0.85	3.62	1.55	0	4.06	1.13	0

Table 1 (cont.)
Item Level Descriptive Statistics

	New Zealand		India		China		United States	
	% ceil	% floor	% ceil	% floor	% ceil	% floor	% ceil	% floor
Phy Concern								
PC1	9.71	1.14	38.14	2.54	21.31	0.82	10.59	0
PC2	5.71	1.14	32.2	2.54	4.92	1.64	12.94	0
PC3	2.86	2.86	13.56	4.24	9.84	0	1.18	2.35
PC4	7.43	0	11.02	1.69	18.85	1.64	4.71	0
PC5	5.71	1.14	17.8	2.54	15.57	0.82	2.35	3.53
Phy View								
PV1	4.57	4	11.86	4.24	3.28	2.46	1.18	1.18
PV2	2.86	4	9.32	5.93	3.28	4.92	1.18	1.18
PV3	1.71	7.43	7.63	13.56	0.82	5.74	1.18	7.06
PV4	2.29	5.14	8.47	5.93	1.64	10.66	1.18	2.35
PV5	3.43	8.57	6.78	17.8	0.82	13.93	4.71	1.18
PV6	2.29	7.43	4.24	12.71	3.28	7.38	1.18	3.53
Achiev Concern								
AC1	6.29	1.71	22.88	11.86	13.93	0	7.06	0
AC2	9.71	2.86	21.19	4.24	16.39	6.56	9.41	0
AC3	8	2.29	35.59	3.39	20.49	3.28	3.53	2.35
AC4	6.86	2.86	33.9	3.39	20.49	0.82	3.53	2.35
AC5	11.43	0.57	33.05	4.24	18.85	1.64	12.94	0
Achiev View								
AV1	1.14	1.14	10.17	4.24	2.46	6.56	4.71	0
AV2	1.14	2.29	9.32	5.08	4.1	5.74	1.18	0
AV3	3.43	1.71	10.17	5.93	4.1	6.56	2.35	0
AV4	1.14	5.14	5.08	13.56	5.74	9.84	1.18	1.18
AV5	2.29	5.14	8.47	11.86	3.28	10.66	1.18	3.53

4.2 Multi-Trait Multi-Item Matrix:

4.2.1 Internal Consistency. As shown in Table 2, across the four samples, the range of correlations of individual items with the target scales are in excess of .4. The diagonal of the multi-trait multi-method matrix provides this information. While it is not shown in Table 2 for the sake of brevity, the only exceptions are items PC1 and PC4 representing physical concern in China and AC5 representing achievement concern in India, where item correlations with their target scales were below .4.

4.2.2 Equality of Item-Scale Correlations. Even though correlations of items with their target scales are not equal as per Table 2, for the most part, items in each sample have contributed significantly to their respective scales (i.e., item-scale correlations for target scales are > .4). Hence, as described in section 3.2.2, it can be concluded that all items representing their respective target scales contribute equally to item-scale correlations.

Table 2A
Multi-Trait Multi-Method Correlation Matrix

	New Zealand				India				China				United States			
	PC	PV	AC	AV	PC	PV	AC	AV	PC	PV	AC	AV	PC	PV	AC	AV
PC	.64-.72	.27-.31	.21-.27	.11-.29	.54-.64	.38-.41	.30-.39	.14-.23	.37-.52	.14-.40	.30-.33	,15-.28	.58-.66	.03-.09	.33-.41	.15-.30
PV	.27-.42	.72-.92	.21-.37	.31-.40	.36-.41	.60-.71	.31-.42	.23-.39	.16-.39	.47-.86	.12-.27	.36-.45	.08-.24	.59-.90	.18-.21	.30-.35
AC	.19-.23	.19-.23	.55-.73	.33-.44	.26-.31	.26-.29	.36-.56	.27-.31	.24-.33	.07-.23	.35-.46	.10-.29	.28-.39	.03-.05	.54-.56	.34-.36
AV	.13-.28	.30-.40	.29-.43	.67-.74	.17-.36	.35-.39	.26-.33	.64-.72	.23-.27	.38-.45	.14-.26	.61-.72	.20-.25	.24-.30	.40-.55	.74-.82

Note: Table shows range of correlations of individual scale items with the scale composite scores of the 4 vanity dimensions – Physical Concern (PC), Physical View (PV), Achievement Concern (AC), and Achievement View (AV). For example, the range of correlations of the 5 PC scale items with the PC scales composite.64 to .72 in New Zealand. The correlations of PC scale items with PV scale in New Zealand range from .27 to .31.

4.2.3 Item External Consistency. Item-scale correlations for non-target scales (i.e., off-diagonal correlations in Table 2) are consistently lower as compared to item-scale correlations for target scales (i.e., correlations that appear on the diagonal). For example, in a general sense, items representing physical concern items have higher correlations with the physical concern dimension than with any other dimension. While not presented in Table 2, even for PC1 and PC4 in China, their correlations with physical concern dimension are higher than their correlations with the other three vanity dimensions. Likewise, AC5 has a higher correlation with its target dimension, achievement view, as compared to its correlation with the other three vanity dimensions.

4.2.3 Item Similarity Coefficients. As shown in Table 3, for all items, the item similarity coefficients, computed using Steenbergen (2014), are higher for target scale dimensions than for non-target scale dimensions. Also, those similarity coefficients are above the recommended benchmark of 0.8 for target scales (cf. Anderson and Gerbing 1982), providing support for uni-dimensionality for each of the vanity scale dimensions.

Table 2B
Item Similarity Coefficients

	New Zealand				India				China				United States			
	PC	PV	AC	AV	PC	PV	AC	AV	PC	PV	AC	AV	PC	PV	AC	AV
PC	.95-.96	.68-.81	.57-.69	.54-.69	.78-.87	.60-.88	.68-.76	.47-.73	.89-.92	.27-.43	.69-.75	.46-.62	.95-.97	.84-.88	.83-.86	.64-.70
PV	.71-.80	.93-.97	.66-.71	.76-.82	.73-.83	.91-.94	.53-.67	.82-.87	.23-.52	.83-.88	.25-.56	.44-.72	.85-.88	.95-.96	.82-.85	.77-.83
AC	.61-.66	.65-.72	.95-.97	.53-.85	.68-.80	.47-.71	.80-.83	.40-.63	.69-.80	.38-.48	.92-.93	.80-.83	.83-.84	.82-.85	.94-.95	.70-.78
AV	.53-.65	.70-.83	.81-.85	.96-.98	.67-.72	.82-.85	.54-.60	.94-.97	.50-.61	.55-.62	.80-.86	.96-.97	.55-.68	.72-.81	.66-.76	.92-.95

Note: Table shows the range of similarity coefficients of individual scale items with the 4 vanity dimensions – Physical Concern (PC), Physical View (PV), Achievement Concern (AC), and Achievement View (AV). For example, the similarity coefficients of the 5 PC items with the PC scale composite range from .95-.96, and with the PV scale those same items' similarity coefficients vary from .68 to .81.

Table 2B
Item Similarity Coefficients

	New Zealand				India				China				United States			
	PC	PV	AC	AV	PC	PV	AC	AV	PC	PV	AC	AV	PC	PV	AC	AV
PC	.95-.96	.68-.81	.57-.69	.54-.69	.78-.87	.60-.88	.68-.76	.47-.73	.89-.92	.27-.43	.69-.75	.46-.62	.95-.97	.84-.88	.83-.86	.64-.70
PV	.71-.80	.93-.97	.66-.71	.76-.82	.73-.83	.91-.94	.53-.67	.82-.87	.23-.52	.83-.88	.25-.56	.44-.72	.85-.88	.95-.96	.82-.85	.77-.83
AC	.61-.66	.65-.72	.95-.97	.53-.85	.68-.80	.47-.71	.80-.83	.40-.63	.69-.80	.38-.48	.92-.93	.80-.83	.83-.84	.82-.85	.94-.95	.70-.78
AV	.53-.65	.70-.83	.81-.85	.96-.98	.67-.72	.82-.85	.54-.60	.94-.97	.50-.61	.55-.62	.80-.86	.96-.97	.55-.68	.72-.81	.66-.76	.92-.95

Note: Table shows the range of similarity coefficients of individual scale items with the 4 vanity dimensions – Physical Concern (PC), Physical View (PV), Achievement Concern (AC), and Achievement View (AV). For example, the similarity coefficients of the 5 PC items with the PC scale composite range from .95-.96, and with the PV scale those same items' similarity coefficients vary from .68 to .81

4.3 Scale level Tests:

Based on Table 4, all four scale dimensions have reasonably high variability across the four samples. Table 5 presents internal consistency estimates of the four vanity scale dimensions on the diagonal. They are all above the recommended value of .7 for acceptable reliability. Also, the reliability estimates are higher than the inter-scale correlations that are shown off the diagonal. For example, in the United States, the reliability estimates of physical concern (0.80) and physical view (0.86) are higher than the correlation between those two measures (0.12), confirming discriminant validity of vanity scale dimensions. Together, these results support internal and external consistency of vanity scale dimensions.

Scale Level Descriptive Statistics

		New Zealand			India		
Scale Statistics							
	Minimum	**Maximum**	**Mean**	**Std. Dev**	**Minimum**	**Maximum**	**Mean**	**Std. Dev**
Phy Concern	8	35	24.19	4.73	5	35	26.84	5.29
Phy View	6	42	22.7	6.35	6	42	24.9	7.66
Achiev Concern	11	37	26.28	4.97	10	39	28.83	6.16
Achiev View	7	37	22.7	4.54	7	37	23.33	6.65
		China				**USA**		
Scale Statistics								
	Minimum	**Maximum**	**Mean**	**Std. Dev**	**Minimum**	**Maximum**	**Mean**	**Std. Dev**
Phy Concern	10	34	24.98	4.59	14	35	23.88	4.27
Phy View	7	39	23.21	6.36	10	41	24.22	5.37
Achiev Concern	11	37	27.45	5.2	14	36	25.81	4.61
Achiev View	7	33	20.63	6.16	13	35	24.71	4.42

Table 5
Reliability Coefficients and Inter-Scale Correlations

Scale reliabilities on diagonal and correlations among subscales									
New Zealand	**phy conc**	**phy view**	**ach conc**	**Ach view**	**China**	**Phy conc**	**Phy view**	**ach conc**	**ach view**
phy conc	0.85				phy conc	0.68			
phy view	0.47	0.91			phy view	0.4	0.79		
ach conc	0.28	0.29	0.83		ach conc	0.38	0.31	0.71	
ach view	0.24	0.46	0.47	0.87	ach view	0.31	0.53	0.22	0.86
India	**phy conc**	**phy view**	**ach conc**	**ach view**	**USA**	**Phy conc**	**Phy view**	**ach conc**	**ach view**
phy conc	0.79				phy conc	0.8			
phy view	0.49	0.87			phy view	0.12	0.86		
ach conc	0.45	0.45	0.76		ach conc	0.45	0.11	0.76	
ach view	0.24	0.48	0.42	0.85	ach view	0.23	0.34	0.53	0.9

Note: Internal consistency estimates of reliability are presented on the diagonal and inter-scale correlations (e.g., correlation of physical concern with physical view) are presented off the diagonal.

In sum, the results of various exploratory tests suggest that for all four dimensions of vanity, the various scaling assumptions have been reasonably met across the four samples. Therefore, each of the four vanity dimensions is uni-dimensional. Items representing individual scale items of each vanity dimension can now be summed (or averaged) to form scale composites for cross-national comparison purposes. The vanity scale has demonstrated cross-national applicability.

5. Discussion:

Cross-national studies that employ likert scale format when measuring consumer behavior constructs must demonstrate that the underlying measures satisfy various assumptions at the item and scale level in order to make them cross-nationally applicable. Tests for assessing scale applicability are typically carried out using SEM. While SEM is preferred for large samples, and also when prior knowledge is available on scale dimensionality of measures, they are not useful when working with small samples. But, small samples are often unavoidable in exploratory cross-national research studies. When faced with small samples, the researchers can still evaluate scaling assumptions by using a battery of descriptive analytics. In this paper we have documented key scaling assumptions and how they can be tested. Performing these tests is critical prior to computing scale composites and making cross-national comparisons of scale mean values. At the end, if the various scaling assumptions are met in individual countries, then the corresponding measure can be applied cross-nationally to examine consumer differences.

References:

Anderson, J.C. & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin, 103 (3), 411-423.

Benson, J. & Fleishman, J.A. (1994). The robustness of maximum likelihood and distribution-free estimation to non-normality in confirmatory factor analysis. Quality and Quantity, 28, 117-136.

Boomsma, A. (1985). Nonconvergence, improper solutions and starting values in LISREL maximum likelihood estimation. Psychometrika, 50, 229-242.

Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix, Psychological Bulletin, 56 (2), 81-105.

Clark, L. A., & Watson, D. (1995). Constructing validity: basic issues in scale development, Psychological Assessment, 7 (3), 309-319.

Dolan, C.V. (1994). Factor analysis of variables with 2, 3, 4, 5 and 7 response categories: a comparison of categorical variables estimators using simulated data. British Journal of Mathematical and Statistical Psychology, 47, 309-326.

Durvasula, S., Netemeyer, R.G., Andrews, J.C., & Lysonski, S. (2006). Examining the cross-national applicability of multi-Item, multidimensional measures using generalizability theory. Journal of International Business Studies, 39 (4), 469-483.

Durvasula, S. and Lysonski, S. (2015). Cross-national applicability of a parsimonious measure of acculturation to global consumer culture. Psychological Reports, 16 (3), 738-750.

Gerbing, D. & Anderson, J.C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. Journal of Marketing Research, 25 (2), 186-192.

Halkias, G., Davvetas, V., & Diamantopoulos, A. (2016). The interplay between country stereotypes and perceived brand globalness/localness as drivers of brand preference, Journal of Business Research 69(9), 3621-3628.

Netemeyer, R. G., W. O. Bearden, & Sharma, S. (2003). Scaling procedures: issues and applications. Thousand Oaks, CA: SAGE.

Netemeyer, R., Durvasula, S. and Lichtenstein, D., (1991), "A Cross-National
Assessment of the Reliability and Validity of the CETSCALE", *Journal of Marketing Research*, 28(August), 320-327.

Netemeyer, R.G., Burton, S., & Lichtenstein, D.R. (1995). Trait aspects of vanity: measurement and relevance to consumer behavior. Journal of Consumer Research, March, 612-625.

Neuberg, S. L., S. G. West, M. M. Thompson, & Judice, T.N. 1997). On dimensionality, discriminant validity, and the role of psychometric analyses in personality theory and measurement: reply to Kruglanski et al.'s (1997) defense of the need for closure scale. Journal of Personality and Social Psychology, 73, 1017-1029.

Nunnally, J.C. & Bernstein, I.H. (1994). Psychometric Theory, New York: McGraw-Hill.

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

Suen, H.K. (1990). Principles of Test Theories. Hillsdale, NJ : Erlbaum.

Saris, W.E. and Hartman, H. (1990). Common factors can always be found but can they also be rejected? Quality and Quantity, 24 (4), 471-490.

Steenbergen, M.R. (2000). Item similarity in scale Analysis. Political Analysis, 8 (3), 261=283.

Ware, John E. Jr. & Gandek, B. (1998). Methods for testing data quality, scaling assumptions, and reliability: The IQOLA project approach," Journal of Clinical Epidemiology, 51 (11), 945-952.