Reliability, Validity and Uni-Dimensionality: A Primer

Pillai N., Vijayamohanan and A., Rjumohan

June 2020

RELIABILITY, VALIDITY AND UNI-DIMENSIONALITY:

A PRIMER

Vijayamohanan Pillai N.

Rjumohan A.

June 2020

----

e-mail: thusspakevm@gmail.com

RELIABILITY, VALIDITY AND UNI-DIMENSIONALITY: A PRIMER

Abstract

Though measurement theory has traditionally got full attention and appreciation in scientific researches, its service in social sciences is young in years. Most of the popular definitions of measurement refer only to measuring objects or events in science; however, the phenomena to be measured in social sciences are mostly abstract ones, unobservable, latent, variables. Measurement in social sciences is hence to seek to construct these latent variables from aggregation of relevant observable items of questions or rating scale items in an interview questionnaire. These constructs are interpreted as manifestations of abstract traits. A number of criteria have come up to test the feasibility of this interpretation such as reliability, validity, uni-dimensionality, etc. The present paper seeks to present a primer on these three criteria.

Reliability, Validity and Uni-Dimensionality: A Primer

There has been considerable demand from researchers for light to be thrown on the problems associated with measurement. Though measurement theory has traditionally got full attention and appreciation in scientific researches, its service in social sciences is young in years. Most of the popular definitions of measurement as “the assignment of numbers to objects or events according to rules” (Stevens 1951: 22), refer only to “objects or events” in science; however, the phenomena to be measured in social sciences are mostly abstract ones, unobservable, latent, variables. Measurement in social sciences is hence to seek to construct these latent variables (hence they are called constructs) from aggregation of relevant observable items of questions or rating scale items in an interview questionnaire (hence they are called scales). These constructs are interpreted as manifestations of abstract traits. A number of criteria have come up to test the feasibility of this interpretation such as reliability, validity, uni-dimensionality, etc. The present paper seeks to present a primer on these three criteria.

Reliability

Fundamentally, *reliability* concerns the extent to which a test or a measuring procedure yields the same results on repeated trials. It goes without saying that when we measure a phenomenon, a certain amount of chance error can creep in the measurement always. In any area of scientific investigation, the goal of error-free measurement though laudable is never attained. Instead, "The amount of chance error may be large or small, but it is universally present to some extent. Two sets of measurements of the same features of the same individuals will never exactly duplicate each other" (Stanley 1971: 356). Thus, *unreliability* is always present to at least a limited extent because repeated measurements never *exactly* equal one another. However, even though repeated measurements of the same phenomenon never precisely duplicate each other, they do tend to be consistent from measurement to measurement. This tendency toward consistency found in repeated measurements of the same phenomenon is referred to as *reliability*. The more consistent the results given by repeated measurements, the higher the reliability of the measuring procedure; conversely the less consistent the results, the lower the reliability.

The internal consistency reliability test is usually assessed using Cronbach's alpha and factor analysis (FA).

Cronbach's alpha is a statistic that measures the reliability in terms of the degree of internal consistency among items on a scale. This measure is a generalization of a coefficient introduced by Kuder and Richardson (1937) to estimate the reliability of scales composed of dichotomously scored items, that is, items scored one or zero depending on whether the respondent does or does not possess the particular characteristic under investigation. According to Nunnally (1978), Cronbach's alpha may be interpreted as the expected correlation between a test and a hypothetical alternative form. It is equal to the reliability only if all of the items are exactly equivalent (Novick and Lewis, 1967). "As a general rule, we believe that reliabilities should not be below 0.80 for widely used scales. At that level, correlations are attenuated very little by random measurement error" (Carmines and Zeller, 1979: 51). However, an alpha coefficient of 0.70 or higher is considered in general acceptable; "in the early stages of predictive or construct validation research," it may be "satisfactory" to "have only modest reliability, e.g., 0.70" (Nunnally and Bernstein, 1994: 264–265). This coefficient is usually written as a function of the number of test items (indicators) and the average inter-correlation among the items. As the number of items increases, Cronbach's alpha also increases. Moreover, keeping the number of items constant, if the average inter-item correlation increases, alpha will also increase.

In the case of the FA, we should check the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO-SMA) and Bartlett's test of sphericity, along with the size of the first eigenvalue and the loadings on the first component. The question behind the FA is "How much collinearity or common variance exists among the variables?" That is, "Is the inter-correlation matrix *factorable*? Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO MSA) are the two tests for this query. The null hypothesis for the Sphericity test is that the inter-correlation matrix is an identity matrix (with unity along the principal diagonal and zeroes as off-diagonal elements, which implies that it comes from a population in which the variables are non-collinear). Rejecting this null with a p-value less than 5% confirms that the data are factorable.

Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (MSA) postulates that if two variables share a common factor with other variables, when the linear effects of the other variables are eliminated, their partial correlation (estimated correlation between the unique factors) will be close to zero (that is, the unique factors are orthogonal, independent). The KMO measure tends to unity, when the variables measure a common factor and to zero when they do not. The following Table illustrates the Kaiser's (1974) criteria:

KMO Value	**Degree of Common Variance**
0.90 to 1.00	Marvellous
0.80 to 0.89	Meritorious
0.70 to 0.79	Middling
0.60 to 0.69	Mediocre
0.50 to 0.59	Miserable
0.00 to 0.49	Don't Factor

In addition to this overall MSA for all the indicators, we can also analyse item-specific MSAs using anti-image correlation matrix (AIC), whose diagonal elements report the MSAs for each individual item. The off-diagonal elements give the negatives of the partial correlations between pairs of items. The correlation matrix is factorable if the MSAs on the diagonal of the AIC are large and the negatives of the partial correlations on the off-diagonal are small (Hair, Anderson, Tatham, and Black, 1995; Tabachnick and Fidell, 2001).

Validity

However, reliability itself is not enough for a measure to provide an accurate representation of some abstract concept. It must also be valid. "In a very general sense, any measuring device is valid if it does what it is intended to do. An indicator of some abstract concept is valid to the extent that it measures what it purports to measure" (Carmines and Zeller, 1979: 12). There are three basic types of validity: content validity, criterion-related validity, and construct validity. Joppe (2000) states that validity in quantitative research determines whether the research truly measures that which it is intended to measure. The validity of a research scale gives us an insight on to how truthful the research results are. In other words the validity can also be described as a soft measure to see whether our research instrument allow us to hit "the bull's eye" of the research objective.

Content validity is achieved when the empirical measure covers the domain of content of the theoretical concept. "Fundamentally, content validity depends on the extent to which an empirical measurement reflects a specific domain of content" (Carmines and Zeller, 1979: 20). However, content validity is taken by necessity as an imprecise standard for evaluating the validity of empirical measurements and has thus limited usefulness (see for details Carmines and Zeller, 1979: Chapter 2).

Along with content validity, another measure used in the social sciences is 'face validity'. Face validation "concerns judgements about an instrument *after* it is constructed" (Nunnally 1978: 111), focusing on the extent to which it "looks like"' it measures what it is intended to measure.

Criterion-related validity "is at issue when the purpose is to use an instrument to estimate some important form of behavior that is external to the measuring instrument itself, the latter being referred to as the criterion" (Nunnally 1978: 87). That is, the degree of criterion-related validity depends on the extent of the correspondence between the test and the criterion. Technically there are two types of criterion-related validity. (i) Concurrent validity is assessed when the criterion exists in the present by correlating a measure and the criterion at the same point in time; whereas (ii) predictive validity concerns a future criterion which is correlated with the relevant measure. "Tests used for selection purposes in different occupations are, by nature, concerned with predictive validity…... Notice that the logic and procedures are the same for both concurrent and predictive validity; the only difference between them concerns the current or future existence of the criterion variable" (Carmines and Zeller, 1979: 18).

Just like the content validity, the criterion-related validity also has only limited generalized applicability in the social sciences, because in many situations there simply do not exist any criteria against which the measure can be reasonably evaluated. Moreover, this difficulty increases along with the degree of abstractness of the concept.

Unlike content validity and criterion-related validity, construct validity is blessed with generalized applicability in the social sciences. We can assess the construct validity of an empirical measure by placing it in theoretical context. Thus, this validation depends on "the extent to which a measure performs in accordance with theoretical expectations. Specifically, if the performance of the measure is consistent with theoretically derived expectations, then it is concluded that the measure is construct valid. On the other hand, if it behaves inconsistently with theoretical expectations, then it is *usually* inferred that the empirical measure does not represent its intended theoretical concept. Instead, it is concluded that the measure lacks construct validity *for that particular concept*" (Carmines and Zeller, 1979: 27).

Construct validity has two sub-categories called convergent validity and discriminant validity. The two measures work together in the sense that if we have evidence for both convergent and discriminant validity, then by definition we have evidence for construct validity. However, one alone is never capable of establishing construct validity.

Convergent validity seeks to show certain measures theoretically supposed to be related to form the same construct are in fact related to each other; that is, it seeks to show a convergence among similar measures. On the other hand, discriminant validity seeks to show two measures that are not theoretically supposed to be related are in fact unrelated; that is, it seeks to discriminate between dissimilar measures.

Since correlation coefficient is used as an estimate of the extent of the relationship between any two measures, we analyse the patterns of inter-correlations among the measures to assess construct validity. It goes without saying that correlations among theoretically similar measures (items or indicators) must be high, whereas those among theoretically dissimilar measures must be low. This is usually carried out in factor analysis by examining the factor-loadings wherein the measure's loading to the corresponding latent construct (factor) should be higher than that on other constructs (Hair et al., 2016). Note that in factor analysis, the items cluster into factors, or more precisely, the items 'load' onto factors; hence, the correlation between an item and its factor is called 'factor loading'. The rule for convergent validity is that the loadings of all measures should be greater than 0.50 (Hair et al., 2010) with an average of greater than 0.7 for every factor. In the case of discriminant validity, we examine the inter-factor correlation matrix; the rule is that the correlations among factors (constructs) should not exceed 0.7, as a correlation greater than 0.7 indicates a higher shared variance (0.7 x 0.7 = 49% shared variance). In this respect it is important to note that in factor analysis, the factors are in fact assumed to be orthogonal to (independent of) each other, suggesting no correlation among them.

It is significant to note that "Internal consistency is a type of convergent validity which seeks to assure there is at least moderate correlation among the indicators for a concept.... Cronbach's alpha is both a validity coefficient and a reliability coefficient..." (Garson 2013: 11-12) Also see Huh, Delorme, and Reid (2006) for an application of validating consumer attitude constructs by means of Cronbach's alpha.

Uni-dimensionality

In cases where one is interested in assessing a latent composite scale (construct) built up from manifest item responses such as of Likert scale, the underlying assumption is that the construct is dominantly uni-dimensional. An essential component of construct validity is assessing the uni-dimensionality of the observed data on item responses, that is, the items seek to measure only one dimension. Dimensionality in general refers to the structure of a specific phenomenon (Pett et al. 2003) and uni-dimensionality thus refers to one dominant phenomenon or latent variable in our case. "A set of items is uni-dimensional if there exists a variable (often called a latent variable, as this variable may not be observed) which 'explains' all the correlations observed between the items" (Bruno 2006: 162). According to Ziegler and Hagemann (2015: 231), "An item is considered uni-dimensional, if the systematic differences within the item variance are only due to one variance source, that is, one latent variable." There are many statistical procedures to examine the dimensionality or the structure of a set of manifest (observed) variables, such as factor analysis or multidimensional scaling. "Ultimately, these procedures ideally obtain an appropriate number of dimensions to justify the use of composite scores and to explain the pattern of correlations among observed variables" (Slocum-Gori and Zumbo 2011: 446). In identifying dimensionality, researchers usually use in factor analysis the eigenvalues-greater-than-one rule, among others.

Reliability, Validity and Uni-dimensionality Checks

We consider a hypothetical study with three constructs (latent variables) with 16 manifest (observed) items (indicators): C1 (with six items), C2 (with five) and C3 (with five indicators) with a small sample size of 30. First we analyse the 16x16 inter-correlation matrix to find the similarity or distance among them. Table 1 clearly supports our expectation of high affinity among the first six items (Q1 to Q6) and their high distance from other items; similarly we find a cluster among the next five items (Q7 to Q11), away from other items, and again another cluster

Table 1: Inter-Correlation Matrix of 16 Indicators

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
Q1	1.000	.722	.861	.873	.722	.796	-.144	-.198	.085	.000	-.198	.055	.136	.027	0.000	.157
Q2	.722	1.000	.722	.614	.700	.666	-.200	-.098	.196	-.050	-.245	-.094	0.000	-.095	-.141	.051
Q3	.861	.722	1.000	.736	.722	.934	-.144	-.198	.085	-.144	-.339	-.082	0.000	-.110	-.136	-.055
Q4	.873	.614	.736	1.000	.614	.665	-.094	-.018	.120	.047	-.157	.196	.267	.144	.134	.259
Q5	.722	.700	.722	.614	1.000	.666	-.200	-.098	.049	-.050	-.245	-.094	0.000	-.095	-.141	.117
Q6	.796	.666	.934	.665	.666	1.000	-.190	-.247	.033	-.190	-.386	-.009	.067	-.050	-.067	-.022
Q7	-.144	-.200	-.144	-.094	-.200	-.190	1.000	.636	.636	.850	.783	-.094	-.283	-.095	-.141	-.037
Q8	-.198	-.098	-.198	-.018	-.098	-.247	.636	1.000	.569	.783	.569	-.018	-.208	-.033	-.069	.234
Q9	.085	.196	.085	.120	.049	.033	.636	.569	1.000	.636	.569	-.018	-.208	-.033	-.069	-.046
Q10	.000	-.050	-.144	.047	-.050	-.190	.850	.783	.636	1.000	.783	-.094	-.283	-.095	-.141	.183
Q11	-.198	-.245	-.339	-.157	-.245	-.386	.783	.569	.569	.783	1.000	.120	-.208	.107	-.069	.191
Q12	.055	-.094	-.082	.196	-.094	-.009	-.094	-.018	-.018	-.094	.120	1.000	.668	.818	.668	.426
Q13	.136	0.000	0.000	.267	0.000	.067	-.283	-.208	-.208	-.283	-.208	.668	1.000	.605	.733	.332
Q14	.027	-.095	-.110	.144	-.095	-.050	-.095	-.033	-.033	-.095	.107	.818	.605	1.000	.740	.420
Q15	0.000	-.141	-.136	.134	-.141	-.067	-.141	-.069	-.069	-.141	-.069	.668	.733	.740	1.000	.394
Q16	.157	.051	-.055	.259	.117	-.022	-.037	.234	-.046	.183	.191	.426	.332	.420	.394	1.000

with four items (Q12 to Q15), far from other items. However, the last item stands alone, aloof from all others, though it is a little closer to the items in the last cluster. Therefore, we tentatively include it in this group, and carry out a factor analysis to assess the factorability of our measurable variables (16 items).

Table 2 reports the overall KMO-MSA as Kaiser's (1974) 'middling' category; the individual item-specific MSAs given along the principal diagonal of Table 3 are mostly larger and the off-diagonal elements, smaller except a few. The Table also shows that the Bartlett's test of

Table 2: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.741
Bartlett's Test of Sphericity	Approx. Chi-Square	379.834
	df	120
	Sig.	0.000

Table 3: Anti-Image Correlation Matrix

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
Q1	.675[a]	-.088	-.520	-.612	-.083	.069	.419	.628	-.008	-.589	-.092	.005	-.056	-.029	-.025	-.084
Q2	-.088	.886[a]	-.132	.021	-.229	.018	.236	.038	-.367	-.099	.066	.082	-.088	-.068	.141	-.062
Q3	-.520	-.132	.697[a]	-.035	-.157	-.717	-.419	-.399	.026	.525	-.005	.083	.151	.022	.015	.153
Q4	-.612	.021	-.035	.824[a]	.030	.133	-.115	-.328	-.010	.139	.127	-.111	-.147	.023	.013	-.067
Q5	-.083	-.229	-.157	.030	.947[a]	.010	.143	-.041	.037	-.073	-.019	.075	-.044	-.034	.079	-.105
Q6	.069	.018	-.717	.133	.010	.812[a]	.032	.193	-.069	-.188	.245	-.171	-.042	-.005	.038	-.128
Q7	.419	.236	-.419	-.115	.143	.032	.679[a]	.281	-.148	-.659	-.329	.061	-.126	-.008	-.058	.198
Q8	.628	.038	-.399	-.328	-.041	.193	.281	.570[a]	-.261	-.698	.198	-.093	-.009	-.040	.073	-.245
Q9	-.008	-.367	.026	-.010	.037	-.069	-.148	-.261	.768[a]	.049	-.285	-.047	.106	.069	-.226	.303
Q10	-.589	-.099	.525	.139	-.073	-.188	-.659	-.698	.049	.600[a]	-.229	.125	.114	.093	-.058	-.045
Q11	-.092	.066	-.005	.127	-.019	.245	-.329	.198	-.285	-.229	.794[a]	-.264	.100	-.168	.287	-.264
Q12	.005	.082	.083	-.111	.075	-.171	.061	-.093	-.047	.125	-.264	.761[a]	-.339	-.547	.025	-.058
Q13	-.056	-.088	.151	-.147	-.044	-.042	-.126	-.009	.106	.114	.100	-.339	.807[a]	.111	-.451	.008
Q14	-.029	-.068	.022	.023	-.034	-.005	-.008	-.040	.069	.093	-.168	-.547	.111	.763[a]	-.458	-.023
Q15	-.025	.141	.015	.013	.079	.038	-.058	.073	-.226	-.058	.287	.025	-.451	-.458	.743[a]	-.207
Q16	-.084	-.062	.153	-.067	-.105	-.128	.198	-.245	.303	-.045	-.264	-.058	.008	-.023	-.207	.697[a]

Note: MSAs are superscripted by 'a' and given along the principal diagonal.

sphericity (379.834) for the correlation matrix (with degrees of freedom of $df = k(k - 1)/2 = 16(16 - 1)/2 = 120$) is highly significant (p given as Sig = 0), leading us to conclude that the inter-correlation matrix is not identity matrix and hence is factorable.

Reliability Tests

Given this light, we now turn to reliability tests of the specific items found to be highly inter-correlated. As already explained, we measure the reliability of variables using Cronbach's internal consistency reliability test; a Cronbach alpha coefficient of 0.7 or higher indicates an acceptable level of reliability (Nunnaly and Berstein, 1994). First we consider the first six items (Q1 to Q6). The Cronbach alpha is found to be 0.943, much higher than the threshold. The scale mean (that is, mean of the construct) is 57.63 with a standard deviation (SD) of 2.619 or variance of 6.861; the small SD indicates that the scores of the respondents do not have wide variations, that they are consistent.

Table 4 reports the relative effect of each item on the overall measures. The first two and the last columns must be self-explanatory. Thus the 'Scale Mean if Item Deleted' column shows the effects on the overall mean of the scale (construct) if an individual item is deleted; for example, if item1 (Q1) is omitted from the final version of the questionnaire, the overall mean of the scale would fall from 57.63 to 48.03. Similar interpretation applies to the second column values. The last column provides the relative significance of each item in the overall reliability: if by removing an item the overall reliability measure increases, we must drop that item from the final questionnaire; on the other hand, if it results in a fall in the overall reliability, the item needs to be retained. In our case, the overall alpha being 0.943, dropping any item would be costlier.

Another measure of relative significance of each item in the total reliability is given by the 'Corrected Item-Total Correlation' column, showing the relationship between the individual item responses and the overall score on the questionnaire. It is expected that a reliable item (question) must have a positive relationship with the overall score, the ideal being above 0.3. If on the other hand an item has a weak positive (less than 0.3) or a negative relationship with the overall, that

Table 4: Item–Total Statistics: First Scale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Q1	48.03	4.654	0.911	0.877	0.922
Q2	47.97	4.999	0.761	0.605	0.940
Q3	48.03	4.654	0.911	0.915	0.922
Q4	48.10	4.852	0.784	0.765	0.938
Q5	47.97	4.999	0.761	0.605	0.940
Q6	48.07	4.754	0.843	0.873	0.931

question fares poorly on reliability and needs to be removed. In our case, all the six items do have much higher relative significance in this sense also.

Finally, we have the 'Squared Multiple Correlation' column, giving us an 'explained' variability measure on a particular item predicted by all the other items under consideration. For example, the square of the multiple correlation of the first item (Q1) with all other five items is given as 0.877. Note that the squared multiple correlation coefficient is the proportion of variance of the dependent variable in a multiple regression that is explained by the independent variables, which is also given by one minus the unexplained proportion of variance. Hence this measure simply says that 87.7 % of the variation in Q1 is explained by all the other five items, a sufficiently high proportion (multiple R^2: Coefficient of determination).

Next we turn to the second cluster of five items. The Cronbach alpha is found to be 0.914, again much higher than the threshold. The scale mean is 23.23 with a standard deviation (SD) of 2.096 (variance of 4.392), indicating consistent scores of the respondents.

Table 5 has the same interpretation as for Table 4 above. The 'Scale Mean if Item Deleted' column shows that if item1 (Q1) is omitted from the final version of the questionnaire, the overall mean of the scale would fall from 23.23 to 18.57. Similarly for other values and for the second column values. The last column shows that the overall alpha being 0.914, dropping any item would lower the overall reliability and hence is not advisable.

Table 5: Item–Total Statistics: Second Scale

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Q7	18.57	2.806	0.844	0.769	0.882
Q8	18.60	2.938	0.722	0.631	0.907
Q9	18.60	3.007	0.673	0.459	0.916
Q10	18.57	2.737	0.898	0.842	0.870
Q11	18.60	2.869	0.773	0.667	0.896

The 'Corrected Item-Total Correlation' column shows that all the five items in our case are positive and much greater than 0.3 and thus have much higher relative significance in this sense also. Finally, the 'Squared Multiple Correlation' column shows that each of the items, except Q9, is closely correlated to other items, or its variance is highly explained by others.

Finally we have the last cluster of five items. The Cronbach alpha is found to be 0.484, much below the threshold. The scale mean is 4.5 with a standard deviation (SD) of 4.369 (variance of 19.086), indicating a little wide variability in the scores of the respondents.

We consider only the last column of Table 6, which is very significant. It shows that the overall alpha being 0.484, dropping any of the first four items would lower the overall reliability and hence is costly, but dropping the last one (Q16) will result in a very high increase in the alpha value.

Table 6: Item–Total Statistics: Third Scale (With All Five Items)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Q12	3.97	16.309	0.615	0.722	0.398
Q13	4.00	16.690	0.515	0.602	0.419
Q14	4.07	16.340	0.612	0.742	0.399
Q15	4.00	16.414	0.586	0.680	0.404
Q16	1.97	3.206	0.445	0.207	0.905

We therefore drop this item (question) from our study, and the new statistics are as follows:

Cronbach's alpha	Scale Mean	Scale SD	Scale Variance
0.905	1.970	1.790	3.206

Table 7 conveys very significant results and is now self-explanatory.

Table 7: Item–Total Statistics: Third Scale (After Dropping Q16)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Q12	1.43	1.840	0.805	0.717	0.871
Q13	1.47	1.913	0.735	0.602	0.896
Q14	1.53	1.844	0.810	0.741	0.870
Q15	1.47	1.844	0.799	0.677	0.873

Table 8 summarises the reliability test results.

Table 8:

Internal Consistency of the Indicators for the Three Constructs

Question (Item) No.	Construct	Cronbach Alpha 1	Cronbach Alpha 2
Q1	C1	0.943	0.943 (No items removed)
Q2			
Q3			
Q4			
Q5			
Q6			
Q7	C2	0.914	0.914 (No items removed
Q8			
Q9			
Q10			
Q11			
Q12	C3	0.484	0.905 Starred item (No. Q16) removed
Q13			
Q14			
Q15			
Q16*			

Validity Check

Thus the reliability analysis helps us weed out the insignificant question (Q16) from the questionnaire; now with the remaining 15 variables, we turn to Factor Analysis to combine these indicators into the expected constructs, the results of which are reported below.

As the inter-correlation matrix is the same as earlier, it is not reproduced here. The SMAs and the sphericity test result are given in Tables 9 and 10 (in terms of anti-image correlations).

Table 9: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.739
Bartlett's Test of Sphericity	Approx. Chi-Square	369.414
	df	105
	Sig.	0.000

Table 9 shows the overall KMO-MSA still as Kaiser's (1974) 'middling' category; and the individual item-specific MSAs given along the principal diagonal of Table 10 are mostly larger and the off-diagonal elements, smaller except a few, as earlier. Table 9 also shows that the Bartlett's test of sphericity (379.834) for the correlation matrix (with degrees of freedom of $k(k - 1)/2 = 15(15 - 1)/2 = 105$) is highly significant (p given as Sig = 0), suggesting that the inter-correlation matrix is factorable.

Table 10: Anti-Image Correlation Matrix

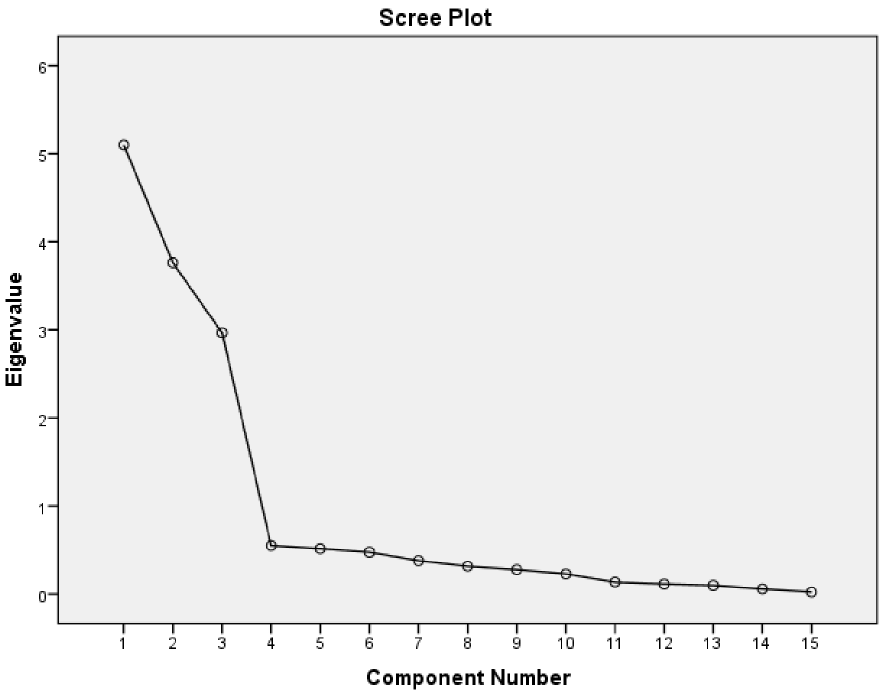
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
Q1	.668[a]	-.094	-.515	-.621	-.092	.059	.446	.629	.018	-.596	-.119	.000	-.056	-.031	-.043
Q2	-.094	.885[a]	-.124	.017	-.238	.011	.253	.023	-.366	-.102	.051	.078	-.088	-.069	.131
Q3	-.515	-.124	.697[a]	-.025	-.143	-.712	-.464	-.378	-.022	.539	.037	.093	.152	.025	.048
Q4	-.621	.017	-.025	.816[a]	.024	.126	-.104	-.356	.011	.136	.113	-.115	-.147	.022	-.001
Q5	-.092	-.238	-.143	.024	.945[a]	-.003	.168	-.069	.073	-.078	-.049	.069	-.044	-.037	.059
Q6	.059	.011	-.712	.126	-.003	.822[a]	.059	.168	-.032	-.195	.221	-.180	-.042	-.008	.012
Q7	.446	.253	-.464	-.104	.168	.059	.662[a]	.346	-.223	-.664	-.293	.075	-.130	-.003	-.018
Q8	.629	.023	-.378	-.356	-.069	.168	.346	.567[a]	-.202	-.732	.143	-.110	-.007	-.047	.023
Q9	.018	-.366	-.022	.011	.073	-.032	-.223	-.202	.824[a]	.066	-.223	-.031	.109	.080	-.175
Q10	-.596	-.102	.539	.136	-.078	-.195	-.664	-.732	.066	.584[a]	-.250	.123	.115	.092	-.069
Q11	-.119	.051	.037	.113	-.049	.221	-.293	.143	-.223	-.250	.830[a]	-.290	.106	-.181	.246
Q12	.000	.078	.093	-.115	.069	-.180	.075	-.110	-.031	.123	-.290	.734[a]	-.340	-.549	.013
Q13	-.056	-.088	.152	-.147	-.044	-.042	-.130	-.007	.109	.115	.106	-.340	.793[a]	.111	-.460
Q14	-.031	-.069	.025	.022	-.037	-.008	-.003	-.047	.080	.092	-.181	-.549	.111	.737[a]	-.474
Q15	-.043	.131	.048	-.001	.059	.012	-.018	.023	-.175	-.069	.246	.013	-.460	-.474	.749[a]

Note: MSAs are superscripted by 'a' and given along the principal diagonal.

We obtain the expected three factors (see Table 11 and the scree plot below) based on the 'Guttman-Kaiser criterion', usually used for determining the appropriate number of significant factors by taking the number of factors with eigenvalues greater than unity (Guttman 1954;

Table 11: Total Variance Explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	5.099	33.991	33.991
2	3.762	25.077	59.068
3	2.966	19.776	78.844
4	.551	3.671	82.515
5	.515	3.435	85.950
6	.476	3.173	89.123
7	.379	2.529	91.652
8	.316	2.108	93.760
9	.278	1.851	95.611
10	.229	1.529	97.140
11	.136	.905	98.045
12	.114	.759	98.803
13	.099	.658	99.461
14	.058	.386	99.847
15	.023	.153	100.000



Kaiser 1960). Note that the three constructs (Table 11) together account for 78.84 % of the total variance of all the 15 variables (items).

The **F**actor Analysis results after varimax rotation (used to generate meaningful factors) are given in Table 12**.**

Table 12: Construct Validity: Factor and Cross Loadings

	Component			Communality
	C1	C2	C3	
Q1	.939	-.022	.081	.889
Q2	.836	-.035	-.100	.710
Q3	.936	-.111	-.095	.898
Q4	.856	.066	.247	.798
Q5	.825	-.070	-.103	.697
Q6	.886	-.171	-.028	.816
Q7	-.128	.890	-.116	.822
Q8	-.110	.813	-.039	.675
Q9	.178	.796	-.041	.666
Q10	-.016	.940	-.106	.895
Q11	-.251	.853	.053	.793
Q12	-.005	.034	.902	.815
Q13	.078	-.238	.831	.754
Q14	-.037	.023	.900	.813
Q15	-.070	-.075	.881	.787

Notes: Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization

a. Rotation converged in 5 iterations

A factor loading represents the degree of correlation between a factor (construct) and a corresponding item (indicator). A higher factor loading (usually greater than 0.5, according to Heir et al., 2006) indicates that an item is closely related to its factor, and a low factor, otherwise. Thus, all the six items, thought to be related to the construct C1 have very high loadings (factor loadings) with it, and very low ones with other factors (cross loadings); similar is the case with other constructs, C2 and C3. Also remember that the six items that constitute C1 are highly correlated among themselves, but less correlated with other factors, as shown also by the inter-correlation matrix, given above. The higher loadings of these inter-correlated C1 items in turn indicate their convergence to the factor C1; and we have convergence validity here. At the same time, the low cross loadings of the C1 items with other factors suggest a sort of discriminant behavior, and thus we have discriminant validity also. Table 13 shows the inter-correlation coefficients among the factors that obey the rule (given above) that the correlations among factors (constructs) should not exceed 0.7, as a correlation greater than 0.7 indicates a higher shared (explained) variance (0.7 x 0.7 = 49% shared variance): another proof for discriminant validity. Note that C1 and C2 are significantly correlated, but their shared variance is only (-0.6 x -0.6 =) 36%.

Table 13: Inter-Correlation Matrix of Constructs (Discriminant Validity)

		C1	C2	C3
C1	Pearson Correlation	1	-.600*	-.390
	Sig. (2-tailed)		.018	.151
	N	15	15	15
C2	Pearson Correlation	-.600*	1	-.444
	Sig. (2-tailed)	.018		.097
	N	15	15	15
C3	Pearson Correlation	-.390	-.444	1
	Sig. (2-tailed)	.151	.097	
	N	15	15	15

*. Correlation is significant at the 0.05 level (2-tailed).

Uni-dimensionality Check

It is significant to note that both the convergence validity and discriminant validity together define uni-dimensionality. All the three constructs turned out to be uni-dimensional. For example, take the case of the first construct C1, composed of six items (indicators) that we found validate both the convergence and discriminant properties. When factor analysis is applied to these six items, we obtain six factors (constructs), but with only the first one having an eigenvalue (4.684) greater than unity, explaining about 78% of the total variance, suggesting only one dominant dimension (latent variable or construct) for these six items; since only one component is extracted, the solution fails to be rotated. The same exercise is applied to the respective items of each of the remaining two constructs and in every case we obtain uni-dimensioanlity, confirming the results on construct validity (see Tables 14, 15 and 16).

Table 14: FA Results for Six Items Q1– Q6 (a)

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.858
Bartlett's Test of Sphericity	Approx. Chi-Square	172.842
	df	15
	Sig.	.000

Table 14: FA Results for Six Items Q1– Q6 (b)

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	4.684	78.062	78.062
2	.468	7.803	85.865
3	.400	6.665	92.530
4	.300	5.000	97.530
5	.093	1.546	99.076
6	.055	.924	100.000

Table 14: FA Results for Six Items Q1– Q6 (c)

	Component
	1
Q1	.941
Q2	.831
Q3	.943
Q4	.851
Q5	.831
Q6	.897

Table 15: FA Results for Five Items Q7– Q11 (a)

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.841
Bartlett's Test of Sphericity	Approx. Chi-Square	104.519
	df	10
	Sig.	.000

Table 15: FA Results for Five Items Q7– Q11 (b)

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	3.741	74.822	74.822
2	.481	9.614	84.436
3	.454	9.071	93.508
4	.213	4.266	97.773
5	.111	2.227	100.000

Table 15: FA Results for Five Items Q7– Q11 (c)

	Component
	1
Q7	.910
Q8	.821
Q9	.779
Q10	.943
Q11	.861

Table 16: FA Results for Four Items Q12– Q15 (a)

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.746
Bartlett's Test of Sphericity	Approx. Chi-Square	76.371
	df	6
	Sig.	.000

Table 16: FA Results for Four Items Q12– Q15 (b)

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	3.118	77.950	77.950
2	.447	11.177	89.127
3	.293	7.313	96.440
4	.142	3.560	100.000

Table 16: FA Results for Four Items Q12– Q15 (c)

	Component
	1
Q12	.895
Q13	.847
Q14	.899
Q15	.890

Also note that we have a few other measures like Fornell-Larcker criterion and Heterotrait-monotrait (HTMT) ratio of correlation for discriminant validity, which are obtainable in the framework of structural equation modelling using Lisrel, AMOS, WarpPLS etc; in this small note we are not considering those measures. In passing, it is significant to remember that "while factor analysis is quite useful for assessing the reliability and validity of empirical measures, it is properly seen as a tool of theoretical analysis, not as a replacement for it" (Carmines and Zeller, 1979: 70).

REFERENCES

Carmines, Edward G. and Zeller, Richard A. (1979) *Reliability and Validity Assessment.* Sage Publications, Inc.

Connelly, L. M. (2008). Pilot studies. Medsurg Nursing, 17(6), 411-2.

Falissard, Bruno (2006) "The unidimensionality of a psychiatric scale: a statistical point of view", *International Journal of Methods in Psychiatric Research*, Volume 8, Number 3: 162-167.

Garson, G. David (2013). *Validity and Reliability* Statistical Associates Publishing, Asheboro, NC 27205 USA.

Guttman, L. (1954). "Some necessary conditions for common factor analysis". *Psychometrika*, Vol. 19: 149-161.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis with readings* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. and Tatham, R. L. (2006). *Multivariate Data Analysis*. 6th edition. Upper Saddle River, New Jersey: Prentice Hall.

Hair, J. F., Money, A. H., Samouel, P. & Page, M. (2007). *Research methods for business*. Chichester: John Willey & Sons Ltd.

Hertzog, M.A. (2008). Considerations in determining sample size for pilot studies. Research in Nursing & Health, 31,180-191.

Hill, R. (1998). What sample size is "enough" in internet survey research? Interpersonal Computing and Technology: An Electronic Journal for the 21st Century, 6(3-4).

Huh, J, Delorme, D. E., and Reid, L. N. (2006). "Perceived third-person effects and consumer attitudes on prevetting and banning DTC advertising". *Journal of Consumer Affairs* 40(1): 90.

Isaac, S., & Michael, W. B. (1995). Handbook in research and evaluation. San Diego, CA: Educational and Industrial Testing Services.

Julious, S. A. (2005). Sample size of 12 per group rule of thumb for a pilot study. Pharmaceutical Statistics, 4, 287-291.

Kaiser, H. F. (1960). "The application of electronic computers to factor analysis". *Educational and Psychological Measurement*, Vol. 20: 141-51.

Kaiser, H. F. (1974). "An index of factorial simplicity". *Psychometrika,*Vol. 39: 32-36.

Kuder, G. F. and M. W. Richardson (1937) "The theory of the estimation of test reliability." *Psychometrika* 2: 151-160.

Novick, M., and Lewis, M. (1967). “Coefficient alpha and the reliability of composite measurements”. *Psychometrika, Vol* 32, 1-13.

Nunnally, J. C. (1978) *Psychometric Theory*. New York: McGraw-Hill.

Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Pett, M. A., Lackey, N. R., and Sullivan, J. J. (2003). *Making sense of factor analysis*. Thousand Oaks, CA: Sage

Sekaran, Uma and Bougie, Roger (2010). *Research methods for business: A skill building approach* (5th ed.). Chichester: John Willey & Sons Ltd

Slocum-Gori, Suzanne L. and Zumbo, Bruno D. (2011). “Assessing the Unidimensionality of Psychological Scales: Using Multiple Criteria from Factor Analysis” *Social Indicators Research*, July, Vol. 102, No. 3, pp. 443-461.

Stanley, J. C. (1971) “Reliability”, pp. 356-442 in R. L. Thorndike (ed.) *Educational Measurement*. Washington, DC: American Council on Education.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Treece, E. W., & Treece, J. W. (1982). Elements of research in nursing (3rd ed.). St. Louis, MO:Mosby.

van Belle, G. (2002). Statistical rules of thumb. New York: John Wiley.

Ziegler, Matthias and Hagemann, Dirk (2015) “Testing the Unidimensionality of Items: Pitfalls and Loopholes”. *European Journal of Psychological Assessment*, Vol. 31(4): 231–237.