Reliability and Validity in Marketing Research: A Statistical Approach

Dr. ASHOKAN.C

Professor & HoD

JBS@JCET

Lakkidi, Ottappalam-679301

Introduction: Marketing research is the process of collecting data about the target market to improve the market share by making suitable modifications in the marketing mix. (4Ps). Usually marketing research will give a lot of valuable information about the underlying motives of consumers and helps in adjusting the promotional policies of the organization. There are various tools to collect and analyze data for this purpose. But the fundamental question to be answered in this case is how reliable and how valid are the data collected for this purpose? Unless the data collected becomes reliable & valid, the whole process of marketing research will become GIGO (Garbage in Garbage Out). None of the latest sophisticated statistical packages and tools can help one replace the authentic data. This article helps to throw light upon the two fundamental concepts of research namely reliability & validity.

Types of Marketing Research: Marketing research falls into two categories; primary & secondary. Primary research focuses on fresh data collected from market to draw conclusions about market whereas secondary research focuses on already collected data used for analysis. In both the cases, statistical analysis is possible to help in arriving at reasonably logical conclusions. But before starting the statistical analysis we have to make sure that the data collected (primary/secondary) is amenable to reliability & validity. Primary data is more suitable for doing a pucca statistical analysis as it is devoid of any first-hand error. Generally secondary data is already baked once and re-baking it may not be that advisable. However, by making suitable adjustments in the data, we can use suitable tools for analysis.

Uses of Marketing Research: Marketing research can be done in five ways: observation, focus groups, surveys, behavioral data, and experiments. Marketing research is extremely useful in demand forecasting, understanding consumer insights, buyer behavior, product positioning and STP. Some times a product-preference test may also become critical in finding out the hidden motives of buyers. Closed ended questions like Dichotomous, Multiple Choice, Likert Scale, Semantic Differential, Importance Scale, Rating Scale & Intention-to-buy scale may reveal a lot of data about the market. Completely unstructured questions like word association, Sentence Completion, Story Completion, Picture & TAT can also be handy in revealing the psychological association of the consumer with the product.

Reliability: Reliability is one of the most basic requirements for a research. Reliability is all about consistency. If a researcher does the experiment 3 times and comes out with 3 different answers then we can say that that research is not reliable. Reliability is all about being uniform in giving results. There are a number of measures of reliability.

Interrater reliability: It is all about giving the instrument to different respondents and measuring the results. Example can be a patient going to a doctor with stomach pain all doctors giving the same diagnosis.

Test-Retest Reliability: The instrument being administered to the same respondent more than once and the result being measured.

Inter-method Reliability: This is a method of finding out the reliability of an instrument by using different methods. Example the weight of an individual can be measured by both normal weighing machine and electronic machine and the difference if any can be considered as a measure of inter-method Reliability.

Internal Consistency: The internal consistency of an instrument can be found by Cronbach's alpha. Alpha is a measure of internal consistency. Usually depending upon the construct, we are measuring, the acceptable value may vary, how-ever any value above 0.6 is considered as an acceptable value to authenticate the consistency of the instrument.

Difference between Reliability & Validity: A reliable measure need not be valid, in validity we are measuring the true value of the construct being measured. A valid measure is always reliable but not vice versa. A wall-clock running 10 minutes fast always is reliable, but not valid. Validity is all about measuring what we intend to measure. Several basic types of validity exist, although often described with somewhat varying terminology. In a highly readable and almost lay-man-like presentation of the subject, Nunally writes of three basic types: (1) content validity which is generally irrelevant in consumer research. (2) predictive validity, (3) construct validity. Face validity, a non-psychometric variety, refers to whether a measure looks like it is measuring what it is supposed to be measuring. Examination of the core consumer behavior journals and conference proceedings since 1970-a body of literature reveals the following:

Face Validity: First, there are numerous examples of face validity. The measures used almost always look like they are measuring that which they are supposed to be measuring. However, the overwhelming majority studies go no further, i.e., provide no empirical support. Thus, face validity is often used as a substitute for construct validity.

Predictive Validity: There are also a sizeable number of studies which suggest the existence of predictive validity, that is, the measure in question seems to correlate, as predicted, with the measure of other variables. Unfortunately, many investigators do not seem to recognize that predictive validity provides little, if any understanding of the relationship. One can have a predictive validity coefficient of 0.99 and still know why or what it means-other than the fact that the scores on one variable are highly predictive of scores on second variable. The relationship may even be meaningless. Obviously high predictive validity doesn't necessarily have to be meaningful.

Cross Validity: One type of predictive validity, however, receives too little attention, namely cross-validity. "Whereas predictive validity is concerned with single sample, cross-validity requires that the effectiveness of the predictor composite be tested on a separate independent sample from the sample population". It should be obvious that unless we can cross-validate our findings, we may really have no findings at all. Again, examination of the literature reveals few cross-validation studies.

Construct Validity: The most necessary type of validity in scientific research is construct validity.

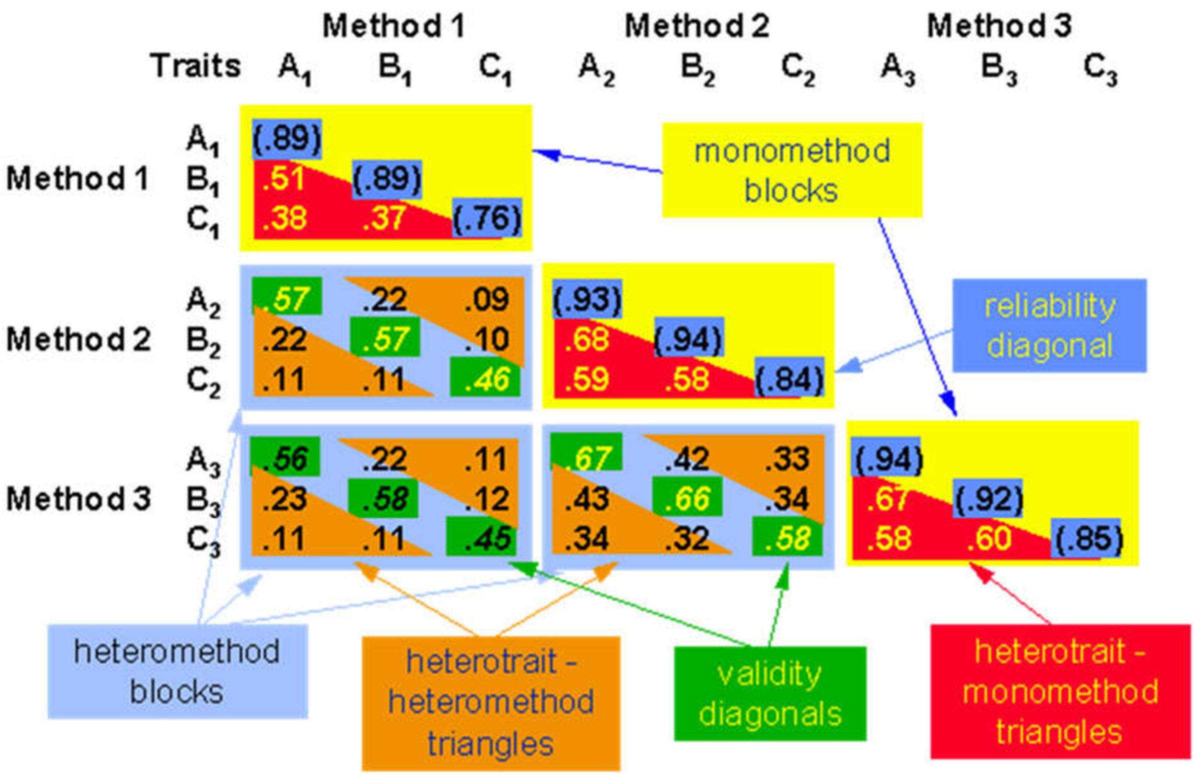
Examination of the recent literature indicates that a negligible proportion of our productivity has been directed toward determining construct validity. A large part of the problem lies in the fact that many researchers appear to be naively believe that scientific research is a game played by creating measures and applying them directly to reality. Although guided by some implicit conceptualization of what it is by trying to measure, the consumer researcher rarely makes his implicit concepts sufficiently or uses them as a basis for developing operational measures. Yet virtually all contemporary scholars of science generally agree that the concept must precede the measure.

Convergent Validity: A basic and relatively easy-to-establish component of construct validity is convergent validity. This refers to the degree to which attempts to measure the same concept using the same concept using two or more different measures yield the same results. Even if few construct validity investigations are available, it seems reasonable to expect that, since many of our core concept are characterized by numerous and varied operationalizations, we should find many studies to demonstrate convergent validity.

The Multitrait-Multimethod Matrix approach for validity

What is the Multitrait-Multimethod Matrix?

The Multitrait-Multimethod Matrix (hereafter labeled MTMM) is an approach to assessing the construct validity of a set of measures in a study. It was developed in 1959 by Campbell and Fiske (Campbell, D. and Fiske, D. (1959). Convergent and discriminant validation by the Multitrait-multimethod matrix. 56, 2, 81-105.) in part as an attempt to provide a practical methodology that researchers could actually use (as opposed to the **nomological network** idea which was theoretically useful but did not include a methodology). Along with the MTMM, Campbell and Fiske introduced two new types of validity – **convergent and discriminant** – as subcategories of **construct validity**. **Convergent validity** is the degree to which concepts that should be related theoretically are interrelated in reality. **Discriminant validity** is the degree to which concepts that should *not* be related theoretically are, in fact, *not* interrelated in reality. You can assess both convergent and discriminant validity using the MTMM. In order to be able to claim that your measures have construct validity, you have to demonstrate both convergence and discrimination.



The MTMM is simply a matrix or table of correlations arranged to facilitate the interpretation of the assessment of construct validity. The MTMM assumes that you measure each of several concepts (called *traits* by Campbell and Fiske) by each of several methods (e.g., a paper-and-pencil test, a direct observation, a performance measure). The MTMM is a very restrictive methodology – ideally you should measure *each* concept by *each* method.

To construct an MTMM, you need to arrange the correlation matrix by concepts within methods. The figure shows an MTMM for three concepts (traits A, B and C) each of which is measured with three different methods (1, 2 and 3) Note that you lay the matrix out in blocks by *method*. Essentially, the MTMM is just a correlation matrix between your measures, with one exception – instead of 1's along the diagonal (as in the typical correlation matrix) we substitute an estimate of the reliability of each measure as the diagonal.

Before you can interpret an MTMM, you have to understand how to identify the different parts of the matrix. First, you should note that the matrix is consists of nothing but correlations. It is a square, symmetric matrix, so we only need to look at half of it (the figure shows the lower triangle). Second, these correlations can be grouped into three kinds of shapes: diagonals, triangles, and blocks. The specific shapes are:

The Reliability Diagonal (monotrait-monomethod)

Estimates of the reliability of each measure in the matrix. You can estimate reliabilities a number of different ways (e.g., test-retest, internal consistency). There are as many correlations in the reliability diagonal as there are measures – in this example there are nine measures and nine reliabilities. The first reliability in the example is the correlation of Trait A, Method 1 with Trait

A, Method 1 (hereafter, I'll abbreviate this relationship A1-A1). Notice that this is essentially the correlation of the measure with itself. In fact such a correlation would always be perfect (i.e., r=1.0). Instead, we substitute an estimate of reliability. You could also consider these values to be monotrait-monomethod correlations.

The Validity Diagonals (monotrait-heteromethod)

Correlations between measures of the same trait measured using different methods. Since the MTMM is organized into method blocks, there is one validity diagonal in each method block. For example, look at the A1-A2 correlation of .57. This is the correlation between two measures of the same trait (A) measured with two different measures (1 and 2). Because the two measures are of the same trait or concept, we would expect them to be strongly correlated. You could also consider these values to be monotrait-heteromethod correlations.

The Heterotrait-Monomethod Triangles

These are the correlations among measures that share the same method of measurement. For instance, A1-B1 = .51 in the upper left Heterotrait-monomethod triangle. Note that what these correlations share is method, not trait or concept. If these correlations are high, it is because measuring different things with the same method results in correlated measures. Or, in more straightforward terms, you've got a strong "methods" factor.

Heterotrait-Heteromethod Triangles

These are correlations that differ in both trait and method. For instance, A1-B2 is .22 in the example. Generally, because these correlations share neither trait nor method we expect them to be the lowest in the matrix.

The Monomethod Blocks

These consist of all of the correlations that share the same method of measurement. There are as many blocks as there are methods of measurement.

The Heteromethod Blocks

These consist of all correlations that do *not* share the same methods. There are (K(K-1))/2 such blocks, where K = the number of methods. In the example, there are 3 methods and so there are (3(3-1))/2 = (3(2))/2 = 6/2 = 3 such blocks.

Principles of Interpretation

Now that you can identify the different parts of the MTMM, you can begin to understand the rules for interpreting it. You should realize that MTMM interpretation requires the researcher to use judgment. Even though some of the principles may be violated in an MTMM, you may still wind up concluding that you have fairly strong construct validity. In other words, you won't necessarily get *perfect* adherence to these principles in applied research settings, even when you do have evidence to support construct validity. To me, interpreting an MTMM is a lot like a physician's reading of an x-ray. A practiced eye can often spot things that the neophyte misses! A researcher who is experienced with MTMM can use it to identify weaknesses in measurement as well as for assessing construct validity.

		P&P			Teacher			Parent		
	Traits	SE_1	SD_1	LC_1	SE_2	SD_2	LC_2	SE_3	SD_3	LC_3
P&P	SE_1	(.89)								
	SD_1	.51	(.89)							
	LC_1	.38	.37	(.76)						
Teacher	SE_2	.57	.22	.09	(.93)					
	SD_2	.22	.57	.10	.68	(.94)				
	LC_2	.11	.11	.46	.59	.58	(.84)			
Parent	SE_3	.56	.22	.11	.67	.42	.33	(.94)		
	SD_3	.23	.58	.12	.43	.66	.34	.67	(.92)	
	LC_3	.11	.11	.45	.34	.32	.58	.58	.60	(.85)

To help make the principles more concrete, let's make the example a bit more realistic. We'll imagine that we are going to conduct a study of sixth grade students and that we want to measure three traits or concepts: Self Esteem (SE), Self Disclosure (SD) and Locus of Control (LC). Furthermore, let's measure each of these three different ways: a Paper-and-Pencil (P&P) measure, a Teacher rating, and a Parent rating. The results are arrayed in the MTMM. As the principles are presented, try to identify the appropriate coefficients in the MTMM and make a judgement yourself about the strength of construct validity claims.

The basic principles or rules for the MTMM are:

Coefficients in the reliability diagonal should consistently be the highest in the matrix. That is, a trait should be more highly correlated with itself than with anything else! This is uniformly true in our example.

Coefficients in the validity diagonals should be significantly different from zero and high enough to warrant further investigation. This is essentially evidence of convergent validity. All of the correlations in our example meet this criterion.

A validity coefficient should be higher than values lying in its column and row in the same heteromethod block. In other words, (SE P&P)-(SE Teacher) should be greater than (SE P&P)-(SD Teacher), (SE P&P)-(LC Teacher), (SE Teacher)-(SD P&P) and (SE Teacher)-(LC P&P). This is true in all cases in our example.

A validity coefficient should be higher than all coefficients in the heterotrait-monomethod triangles. This essentially emphasizes that trait factors should be stronger than methods factors. Note that this is *not* true in all cases in our example. For instance, the (LC P&P)-(LC Teacher) correlation of .46 is less than (SE Teacher)-(SD Teacher), (SE Teacher)-(LC Teacher), and (SD

Teacher)-(LC Teacher) – evidence that there might me a methods factor, especially on the Teacher observation method.

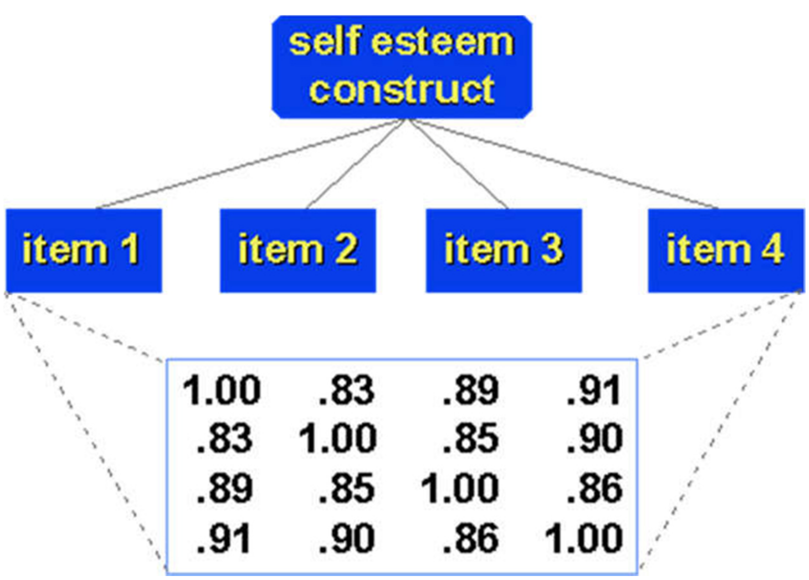
The same *pattern* of trait interrelationship should be seen in all triangles. The example clearly meets this criterion. Notice that in all triangles the SE-SD relationship is approximately twice as large as the relationships that involve LC.

Advantages and Disadvantages of MTMM

The MTMM idea provided an operational methodology for assessing construct validity. In the one matrix it was possible to examine both convergent and discriminant validity simultaneously. By its inclusion of methods on an equal footing with traits, Campbell and Fiske stressed the importance of looking for the effects of how we measure in addition to what we measure. And, MTMM provided a rigorous framework for assessing construct validity.

Despite these advantages, MTMM has received little use since its introduction in 1959. There are several reasons. First, in its purest form, MTMM requires that you have a fully-crossed measurement design – each of several traits is measured by each of several methods. While Campbell and Fiske explicitly recognized that one could have an incomplete design, they stressed the importance of multiple replication of the same trait across method. In some applied research contexts, it just isn't possible to measure all traits with all desired methods (would you use an "observation" of weight?). In most applied social research, it just wasn't feasible to make methods an explicit part of the research design. Second, the judgmental nature of the MTMM may have worked against its wider adoption (although it should actually be perceived as a strength). many researchers wanted a test for construct validity that would result in a single statistical coefficient that could be tested – the equivalent of a reliability coefficient. It was impossible with MTMM to quantify the *degree* of construct validity in a study. Finally, the judgmental nature of MTMM meant that different researchers could legitimately arrive at different conclusions.

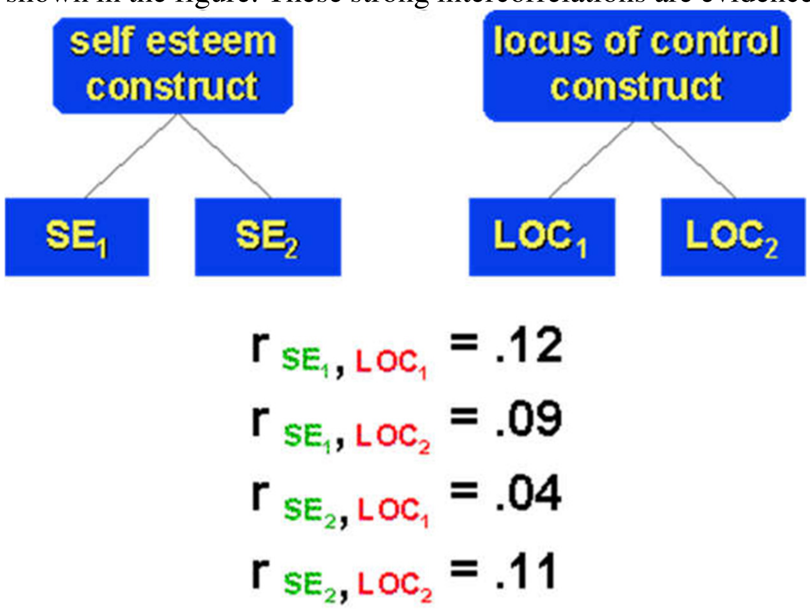
A Modified MTMM – Leaving out the Methods Factor



As mentioned above, one of the most difficult aspects of MTMM from an implementation point of view is that it required a design that included all combinations of both traits and methods. But the ideas of convergent and discriminant validity do not require the methods factor. To see this, we have to reconsider what Campbell and Fiske meant by convergent and discriminant validity.

What is convergent validity?

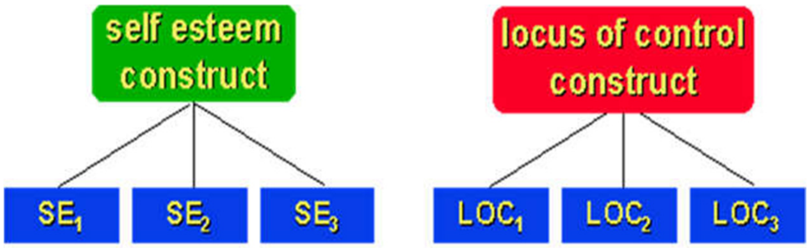
It is the principle that *measures of theoretically similar constructs should be highly intercorrelated*. We can extend this idea further by thinking of a measure that has multiple items, for instance, a four-item scale designed to measure self-esteem. If each of the items actually does reflect the construct of self-esteem, then we would expect the items to be highly intercorrelated as shown in the figure. These strong intercorrelations are evidence in support of convergent validity.



And what is discriminant validity?

It is the principle that *measures of theoretically different constructs should not correlate highly with each other*. We can see that in the example that shows two constructs – self-esteem and locus of control – each measured in two instruments. We would expect that, because these are measures of different constructs, the cross-construct correlations would be low, as shown in the figure. These low correlations are evidence for validity. Finally, we can put this all together to see how we can address both convergent and discriminant validity simultaneously. Here, we have two constructs – self-esteem and locus of control – each measured with three instruments. The red and green correlations are within-construct ones. They are a reflection of convergent validity and should be strong. The blue correlations are cross-construct and reflect discriminant validity. They should be uniformly lower than the convergent coefficients.

The important thing to notice about this matrix is that *it does not explicitly include a methods factor* as a true MTMM would. The matrix examines both convergent and discriminant validity (like the MTMM) but it only explicitly looks at construct intra- and interrelationships. We can see in this example that the MTMM idea really had two major themes. The first was the idea of looking simultaneously at the pattern of convergence and discrimination. This idea is similar in purpose to the notions implicit in the **nomological network** – we are looking at the pattern of interrelationships based upon our theory of the nomological net. The second idea in MTMM was the emphasis on methods as a potential **confounding factor**.



	SE_1	SE_2	SE_3	LOC_1	LOC_2	LOC_3
SE_1	1.00	.83	.89	.02	.12	.09
SE_2	.83	1.00	.85	.05	.11	.03
SE_3	.89	.85	1.00	.04	.00	.06
LOC_1	.02	.05	.04	1.00	.84	.93
LOC_2	.12	.11	.00	.84	1.00	.91
LOC_3	.09	.03	.06	.93	.91	1.00

While methods may confound the results, they won't necessarily do so in any given study. And, while we need to examine our results for the potential for methods factors, it may be that combining this desire to assess the confound with the need to assess construct validity is more than one methodology can feasibly handle. Perhaps if we split the two agendas, we will find that the possibility that we can examine convergent and discriminant validity is greater. But what do we do about methods factors? One way to deal with them is through replication of research projects, rather than trying to incorporate a methods test into a single research study. Thus, if we find a particular outcome in a study using several measures, we might see if that same outcome is obtained when we replicate the study using different measures and methods of measurement for the same constructs. The methods issue is considered more as an issue of generalizability (across measurement methods) rather than one of construct validity.

When viewed this way, we have moved from the idea of a MTMM to that of the Multitrait matrix that enables us to examine **convergent and discriminant validity**, and hence construct validity. We will see that when we move away from the explicit consideration of methods and when we begin to see convergence and discrimination as differences of degree, we essentially have the foundation for the **pattern matching** approach to assessing construct validity.

Reference:

1. . Nunnally J, Bernstein L. Psychometric theory. New York: McGraw-Hill Higher, INC; 1994.

2. Cortina J.What is coefficient alpha: an examination of theory and applications. *Journal of applied psychology.* 1993; 78:98-104. 10.1037/0021-9010.78.1.98 [CrossRef] [Google Scholar]
3. Cronbach L.Coefficient alpha and the internal structure of tests. *Psychomerika.* 1951; 16:297-334. 10.1007/BF02310555 [CrossRef] [Google Scholar]
4. Campbell DT and Fiske DW. Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin* 1959; 56:81-105