

COMMON RESEARCH TERMS, DEFINITIONS, AND USES



Compiled and Prepared by
Long Bunteng
For Post-Graduate Studies

Table of Contents:

1. Inductive Versus Deductive Researches	3
2. Survey Research and where it is used	4
3. The difference between a Hypothesis and a Theory	6
4. Difference between Applied and Fundamental Research	7
5. Sample Size	8
6. Statistics versus Parameter	9
7. Population versus Sample	10
8. P values	11
9. KMO and Bartlett's test of sphericity	14
10. KMO	15
11. Bartlett's test of Sphericity	16
12. Measurement versus Residual Error Terms	17
13. Exogenous versus Endogenous variables	18
14. Communalities	19
15. The meaning of R, R Square, Adjusted R Square, R Square Change and F Change in a regression analysis	21
16. Discriminant Validity through Fronell-Larcker Criterion	29
17. Discriminant Validity	33
18. Confirmatory Factor Analysis (CFA)	38
19. Convergent Validity	39
20. Factor loading and Cross-loading	41
21. How to calculate Average Variance Extracted and Composite Reliability	55
22. Discriminant Validity through Cross Loadings	58
23. Measurement versus Structural Model	60
24. Average Variance Extracted (AVE)	62
25. Composite Reliability	64
26. Latent versus Observed variables	65
27. Exploratory Factor Analysis (EFA)	67

Inductive versus deductive research

Inductive research/observation/reasoning means using a particular observation to generate general explanation. For example, after eating sandwich from KFC over a month, you concluded that all sandwich at KFC outlets are delicious. i.e. *from specific to general*.

Deductive research/observation/reasoning means predicting a particular outcome based on a general observation. Considering the KFC example, if after tasting the Sandwich sold in all KFC outlet in a certain small town, you came to the conclusion that the Sandwich from one particular outlet (say outlet G) is the most delicious. i.e. *from general to specific*.

Survey Research and where it is used

In most works of literature, survey has been defined as the collection of facts and figures to gain insight about a specific subject, hence quantitative in nature.

Survey can also be qualitative especially when they are used to collect qualitative data. For example, when a researcher collects data from respondents using an open-ended questionnaire, they are simply gathering qualitative data which is very rich in information.

This post focuses on survey research that are objective and information are gathered using a probabilistic sampling method where each respondent has a non-zero chance of participating in the study. The good thing about this approach is that it helps researchers to generalize their findings to the entire population.

Where is it used:

Survey can be used in various fields, for instance

- companies can use survey to collect data about a particular product so as to enhance their

marketing strategies and enhance consumer preferences and interests;

- politicians use opinion polls to measure the political perception and opinions about their citizens;
- in academic discipline, surveys can be used to collect data on respondent's attitude, beliefs, motivation, emotions etc. about a phenomenon.
- the Bureau of Labor Statistics conduct survey to collect data on people's income, the availability of jobs, and unemployment rates.
- the record department of the police service uses survey to gather data about criminal events and measure people's fear and concerns about crime.

The difference between a Hypothesis and a Theory

Hypothesis are claims about observations that can either be accurate or inaccurate.

For example, the statement: the average weight of Americans studying at Harvard is 45kg. This statement can either be true or false. We can only test this hypothesis by recording the weight of each and every American student studying at Harvard and then finding out whether the average weight is indeed 45kg.

A theory, on the other hand, are hypotheses that have been empirically tested to be true. Theories are considered highly plausible because they have withstood attempt to refute on logical grounds, historical, qualitative or quantitatively analysis.

Difference between Applied and Fundamental Research

Applied research are those research that are meant to directly solve problems to improve human conditions.

An example is the research conducted to produce vaccine for the deadly coronavirus. Why? Because scientist want to eradicate the virus from the earth's surface, hence saving lives.

Fundamental research is a kind of research solely meant for knowledge acquisition. It helps to expand our knowledge beyond our immediate surroundings. This kind of research rarely help to improve human lives.

For example, we might be interested in investigating the relationship between happiness and Facebook usage i.e. being happy means you use Facebook. The idea is to ONLY know whether or not this relationship is the same across different countries, race, ethnicities, gender etc. Note that the key here is not HOW happiness leads to Facebook usage.

Sample Size

Sample size plays a very important role in every research you undertake. That is why it is very essential you gather enough information about your subject. In statistics, a sample size of at least 30 is considered large enough, however, the larger the sample size, the more accurate information you can obtain and generalize your result to the entire population.

In other words, the larger the sample size, the closer your **sample statistic** will be to the actual **population parameter** and the least error you can recall in the study. Also, remember that the sample size obeys the law of diminishing returns. i.e. as you increase the sample size, it gets to a point where successive increase doesn't really add any effect but rather reduce the accuracy of your output. That is why you need to ensure that your sample mean approximate the population mean, hence exhibiting the properties of a normal distribution.

Statistics versus Parameter

Let's assume from the scenario illustrated here, that the average income of all the customers is \$1,300/month. Given that this amount is calculated from the sample drawn from the population, it is called a **Statistic** or **Sample Statistic**.

If you follow the assumptions rightly, then we can generalize this result to the whole population, thus the average income of all the customers in the country is \$1,300.00 and this is known as the **Parameter** or the **Population Parameter**. In statistics, it is rare to obtain the parameter from a real population in a study. Hence, we use the statistic to make inferences about the general population.

Population versus Sample

Imagine this, you (as a researcher) have been tasked by a large marketing company in your country to find the average income of all their customers to help them formulate new marketing strategies to enrich consumer preference and interests. The right way of doing this is to collect data on all their income and then divide it by the total number of customers.

This looks impossible right? Yes, especially when data must be obtained from over millions of customers scattered all over the country. Since you can't gather data on all these customers, it is advisable you collect data from a subset of the customers. This subset is known as a **sample** and all the customers in the country are called **Population**. The researcher must carefully select this sample using some assumptions:

- the sample selected is directly from the target population (customers of the marketing company)
- the customers are selected randomly
- a valid sample size is selected so as to make generalization to the entire population of customers

P values

P-value is also known as probability value and it ranges from 0 to 1. The interpretation of a p-value varies depending on what one is studying. For instance, while a small p-value less than 0.05 provide good evidence against the null hypotheses for ANOVA, t-test, Pearson's correlation, regression analysis, etc., in some cases such as the statistical test for normality assumption (Kolmogorov-Smirnov and Shapiro-Wilk test), a p-value greater than 0.05 is a good indication that the data is normally distributed (i.e. higher level of significance). This p-value we are talking about is present in any statistical software you can think about.

Quick Facts

A smaller p-value provides greater statistical compatibility of the data with the proposed research hypothesis (alternative hypothesis) given that the assumptions underlying p-value are valid. Hence, providing enough evidence in favour of your hypothesis. P-value has nothing to do with the hypothesis you are testing or observing (alternative hypothesis). In other words, it does not tell you whether or not a given hypothesis is true. Instead, it gives you information about the odds of observing that hypothesis (null hypothesis). For example, a p-

value of 0.01 means that there is a 1% chance that the null hypothesis will be rejected incorrectly and 99% evidence in support of the alternative hypothesis.

A p-value does not tell us the relative importance of a variable neither does it provide information on the size of an effect on a variable. For example, consider the variables A and B. Variable A has a p-value of 0.002 and variable B has a p-value of 0.03. The fact that the p-value of A is smaller than B doesn't necessarily mean A is more important than B neither does it mean A exert a larger effect than B. You see, the reason why you need to take this seriously is that your sample size and the accuracy of your measurement instruments influence the p-value.

Always remember that p-value is not the only way to conclude your study or research. You can as well solidify your research findings by using predictive or confidence intervals, likelihood ratio, false discovery rate, etc.

Let's understand this using a practical example:

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.137	.226		18.329	.000
	Political ties	-.064	.059	-.060	-1.085	.279
	Business ties	.136	.040	.187	3.415	.001

a. Dependent Variable: Firm Performance

In the SPSS output above, the p-value is listed under the column Sig. We can see that the p-value for *business ties* is 0.01 which is less than 0.05 (a common threshold, you can choose your threshold, for example, 0.10 or 10%) and very significant. This means that there is a 0.01 or 1% chance (a very small chance) that the result on business ties (data on business ties) occurred by chance and 99% this result did not occur by chance.

In the same way, you can see the p-value on *political ties* is .123 which is greater than 0.05, hence not significant. In other words, there is a .123 or 12.3% chance the results of political ties occurred by chance exceeding the 5% threshold. Thus, we reject that hypothesis and conclude the political ties doesn't influence firm performance positively.

KMO and Bartlett's test of sphericity

The table below presents two different tests: the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's test of Sphericity.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.853
Bartlett's Test of Sphericity	Approx. Chi-Square	7413.928
	df	153
	Sig.	.000

KMO

KMO is a test conducted to examine the strength of the partial correlation (how the factors explain each other) between the variables. KMO values closer to 1.0 are consider ideal while values less than 0.5 are unacceptable. Recently,most scholars argue that a KMO of at least 0.80 are good enough for factor analysis to commence. Below is a tabular chart for your perusal.

KMO measure	Interpretation
$KMO \geq 0.90$	Marvelous
$0.80 \leq KMO < 0.90$	Meritorious
$0.70 \leq KMO < 0.80$	Average
$0.60 \leq KMO < 0.70$	Mediocre
$0.50 \leq KMO < 0.60$	Terrible
$KMO < 0.50$	Unacceptable

From our result, we had a KMO value of .853. This indicates that the degree of information among the variables overlap greatly/the presence of a strong partial correlation. Hence, it is plausible to conduct factor analysis.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.853
Bartlett's Test of Sphericity	Approx. Chi-Square	7413.928
	df	153
	Sig.	.000

Bartlett's test of Sphericity

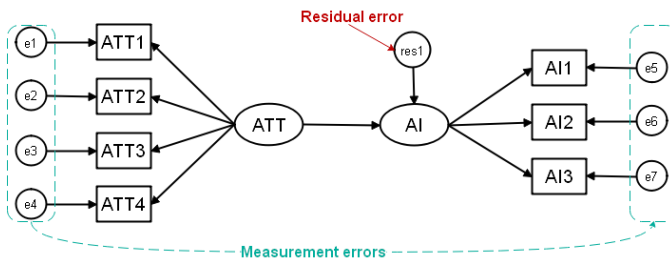
The Bartlett's test of Sphericity is used to test the null hypothesis that the correlation matrix is an identity matrix. An identity correlation matrix means your variables are unrelated and not ideal for factor analysis. A significant statistical test (usually less than 0.05) shows that the correlation matrix is indeed not an identity matrix (rejection of the null hypothesis) as represented in the table below.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.853
Bartlett's Test of Sphericity	Approx. Chi-Square	7413.928
	df	153
	Sig.	.000

Measurement versus Residual Error Terms

In Structural Equation Modeling (SEM), measurement errors are those errors that occur in your items/questions in a questionnaire while measuring a latent variable. Each of these items is assigned a unique item error/measurement error to take care of the flaws or discrepancies that may have arisen in your measurement instruments.

Residual Errors also called disturbance terms are the errors that arise in the prediction of an endogenous variable by an exogenous variable. Below is a diagrammatical illustration of measurement and residual error terms in a SEM model.



Exogenous versus Endogenous variables

Exogenous variables are also known as predictors/independent variables. They are those variables that cause changes in other variables (endogenous variable). Changes that occur in the exogenous variables are as a result of respondents' demographical characteristics and hence cannot be explained by the model.

Endogenous variables, on the other hand, are synonymous with the outcome/dependent variables. Any changes that occur in the endogenous variable are either directly or indirectly caused by the exogenous variables.

Communalities

When we talk about communalities, we are interested in the amount of original information contained in each variable that can be extracted from a common factor. That is, the higher the communality (closer to 1) the larger the amount of information that will be extracted. A good way to measure factor analysis is to have a communality of at least 0.40. By default, the initial value of communality for each variable is 1.00 or 100% because at the beginning of the factor analysis, none of this information has been extracted.

From the result, the minimum amount of information is 56.6% by PEU3 (It is easy to add or edit information in Wikipedia) indicating 43.4% loss of information.

Communalities

	Initial	Extraction
Qu1	1.000	.798
Qu2	1.000	.748
Qu3	1.000	.652
Qu5	1.000	.600
Use1	1.000	.681
Use2	1.000	.637
Use3	1.000	.777
Use4	1.000	.740
SA1	1.000	.747
SA2	1.000	.726
SA3	1.000	.749
Inc1	1.000	.652
Inc2	1.000	.665
Inc3	1.000	.759
Inc4	1.000	.628
PEU1	1.000	.622
PEU2	1.000	.639
PEU3	1.000	.566

Extraction Method: Principal
Component Analysis.

The meaning of R, R Square, Adjusted R Square, R Square Change and F Change in a regression analysis

Concepts can be easily understood when they are explained in a non-technical way with good examples. We explain what these terms mean using a very simple example.

A researcher is interested in examining university faculty user behaviour of information in Wikipedia as a teaching resource using variables such as Quality, Perceived Ease of Use, and Sharing Attitude of faculty members towards information in Wikipedia. Using these variables, the researcher conducted a multiple linear regression and the output summary is shown in the table below. Variables in the model include:

Independent Variables (Predictors):

- Quality of information in Wikipedia
- Sharing Attitude of faculty members and
- Perceived ease of use of information in Wikipedia

Dependent Variable (Outcome): University faculty member's User Behaviour

Table 1: Model Summary^d

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			
						F Change	df1	df2	Sig. F Change
1	.520 ^a	.270	.269	.81387	.270	337.255	1	911	.000
2	.525 ^b	.276	.274	.81111	.006	7.206	1	910	.007
3	.525 ^c	.276	.274	.81143	.000	.277	1	909	.599

a. Predictors: (Constant), Quality

b. Predictors: (Constant), Quality, Sharing Attitude

c. Predictors: (Constant), Quality, Sharing Attitude, Perceived Ease Use

R

R in a regression analysis is called the correlation coefficient and it is defined as the correlation or relationship between an independent and a dependent variable. It ranges from -1 to +1. An R-value of -1 and +1 indicates respectively a perfect negative and positive relationship between the independent and dependent variable. Thus, an R-value of 0 shows that there is no relationship between these variables. So, depending on your study, the higher the R-value, i.e. closer to -1 or +1, the better the relationship. Most often, it is expressed in percentages.

From table 1, in the first Model, there is a positive relationship between the Quality of information in Wikipedia and the University faculty member's User Behaviour. In other words, there exist a 0.346 (34.6%) chance that the Quality of information in Wikipedia will enhance the user behaviour of University faculty members.

In the Second Model, both Quality of information and Sharing Attitude have a positive relationship with User behaviour. Their combined effect has a 0.374 (37.4%) chance of increasing User behaviour.

In the final block, there is a 41.9% chance that all the predictor variable increases user behaviour of University faculty members.

R square

It is also known as the coefficient of determination explains the variations in the dependent variable accounted for by the independent variable. It ranges from 0 to 1 but typically expressed as a percentage during interpretation. It is obtained by squaring the R-value described above. In our example,

If $R = .520$; then R square

Also if $R = .525$; then R square

So, in model 1, an R square of .270 means 27.0% of the variation or change in User Behaviour is accounted for by the Quality of information in Wikipedia. Similarly, in model 2, 27.6% of the changes in User Behaviour is accounted for by both Quality of information in Wikipedia and the Sharing Attitude of University faculty members.

Finally, we can see that 27.6% of the changes in User Behaviour is accounted for by all the predictor variables in our model. In other words, the R square value is not different from the one in model 2, so Perceived ease of use added no change to the study model.

Since this is about human behaviour and because there are several uncontrollable factors influencing User Behaviour, it is reasonable for the predictor variables to have a 27.6% change in User Behaviour.

Adjusted R Square

Assuming you need a higher R square value, you can simply increase the number of independent variables in your model. In other words, R square increases with an increase in the number of independent variables. To curb this situation, an adjusted R square was introduced. Adjusted R square, as the name implies, adjusts the number of independent variables in the model and only improves when the new variable added improves the model; decreases when the new variable does not affect the model. Most often, adjusted r square is reported for a sufficiently complex model with a lot of predictors.

In our example, each of the predictors added with the exception of Perceived ease of use improved the model, hence the adjusted r square increased. Thus, in our case, 27.6% (see block 2 or 3) of the variations in User Behaviour is explained by Quality of information in Wikipedia and the Sharing Attitude of university faculty members.

#R Square Change You see, as you keep adding more variables to your model in what we called hierarchical regression analysis, you may be interested in finding out the contribution of the new variable in the model.

For example, when Sharing Attitude was added to Quality in model 2, the change in r square was 0.006 meaning Sharing Attitude contributed 0.6% of the variation in User Behaviour.

F Change

An F change is a test based on F-test used to determine the significance of an R square change. A significant F change implies the variable added significantly improves the model prediction.

From this explanation, it is easy to see that each of the variables added in block 2 (Sharing Attitude) is significant at a p-value of 0.007. Also, the variable added in block 3 (Perceived ease of use) is not significant since the p-value is greater than 0.05. In other words, only one of these new predictors influences User Behaviour.

Discriminant Validity through Fronell- Larcker Criterion

Before reading this post, we assume the reader has at least a basic understanding of discriminant validity as explained here.

The Fronell-Larcker criterion is one of the most popular techniques used to check the discriminant validity of measurements models. According to this criterion, the square root of the average variance extracted by a construct must be greater than the correlation between the construct and any other construct. Once this condition is satisfied, discriminant validity is established.

Below are the steps necessary for establishing the Fronell-Larcker criterion.

1. Compute the Average for each construct as demonstrated here
2. Find the correlation among the construct (you must use the average you computed for each construct in step 1 above).

Don't know how to find correlation in SPSS, check here.

The SPSS output is shown below.

		Qu_av	Use_av	SA_av	Inc_av
Qu_av	Pearson Correlation	1	.520**	.312**	.232**
	Sig. (2-tailed)		.000	.000	.000
	N	913	913	913	913
Use_av	Pearson Correlation	.520**	1	.234**	.236**
	Sig. (2-tailed)	.000		.000	.000
	N	913	913	913	913
SA_av	Pearson Correlation	.312**	.234**	1	.270**
	Sig. (2-tailed)	.000	.000		.000
	N	913	913	913	913
Inc_av	Pearson Correlation	.232**	.236**	.270**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	913	913	913	913

** . Correlation is significant at the 0.01 level (2-tailed).

Table 1

Table 1 presents you with Pearson Correlation, level of significance of this correlation (Sig. 2 tailed) and the sample size (N). We are interested in the Pearson correlation together with the asterisk (*) indicating the significance level.

Construct	Qu	Use	SA	Inc
Qu	1	.520**	.312**	.232**
Use	.520**	1	.234**	.236**
SA	.312**	.234**	1	.270**
Inc	.232**	.236**	.270**	1

Table 2

So, we create a new table (see Table 2) and input these items. Note that for correlations, the items on the upper diagonals are the same as those on the lower diagonals (see Table 2). Hence, we delete the upper diagonals (see Table 3).

Construct	Qu	Use	SA	Inc
Qu	1			
Use	.520**	1		
SA	.312**	.234**	1	
Inc	.232**	.236**	.270**	1

Table 3

Nice! But we are not done yet.

Next, we replace the ones (“1”) on the diagonal with the square roots of its AVE. For example, the correlation between SAT and SAT is 1 .i.e. $\text{corr}(\text{Qu}, \text{Qu}) = 1$, so we replace the 1 with $\sqrt{\text{AVE for Qu}} = \sqrt{0.658} = 0.811$

Similarly, for Use,

etc.

So, you need to do these for all the constructs and record your values.

Construct	Qu	Use	SA	Inc
Qu	.811			
Use	.520**	.805		
SA	.312**	.234**	.863	
Inc	.232**	.236**	.270**	.817

Table 4

Once this is done, you need to check and make sure the correlation of a construct with other constructs are less than the square root of its AVE.

For example,

1. $\text{corr}(\text{Use}, \text{Qu}) = 0.520 < 0.805$ (True)
2. $\text{corr}(\text{SA}, \text{Qu}) = 0.312 < 0.863$ and
 $\text{corr}(\text{SA}, \text{Use}) = 0.234 < 0.863$ (True)

Discriminant Validity

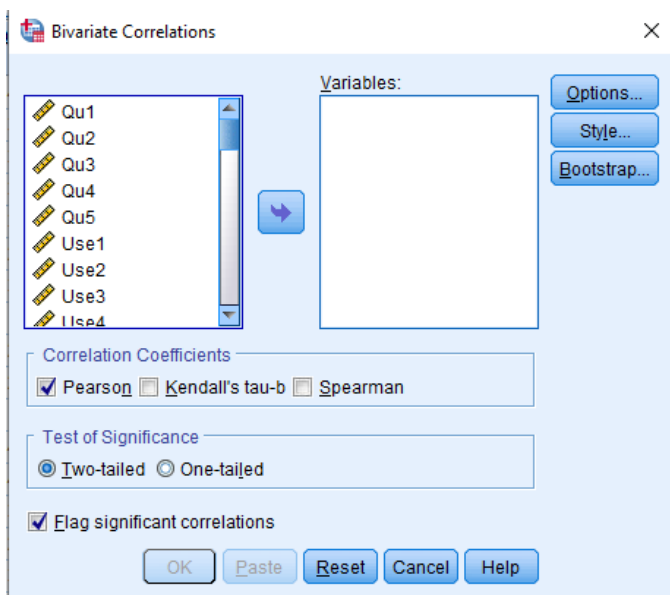
Your main reason for conducting discriminant validity for your study will be to show how distinct an item or set of items is from others. In other words, you are interested in showing that items measuring different constructs or variables have poor relationships or low correlation exist between them. If you fail to test for discriminant validity among your variables, there could be a chance that some items will have a good relationship with non-related items – a clear indication of a lack of discriminant validity.

Let's take a look at this example:

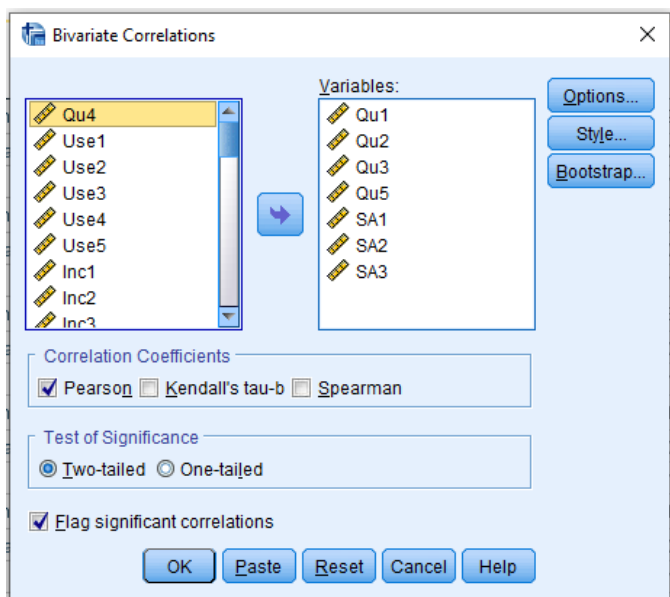
Suppose we have two different constructs, Perceived Quality of Information in Wikipedia (Qu1, Qu2, Qu3 and Qu5) and Sharing Attitude (SA1 – SA3), then as shown in Table 2, the bolded items in red represents discriminant validity. As you can see, the correlation between SA2 and Q1 is **.239**, which is lower when compared to the correlation between items of the same construct like SA3 and SA2 (**.616**) or Qu2 and Qu1 (**.723**).

The easiest way to establish discriminant validity is through correlation coefficient. In SPSS, we can achieve this purpose by following the steps below:

1. Click Analysis → Correlate → Bivariate.
You will be presented with the window below.



2. Move all the items measuring the constructs into the Variable: box. In our case, Perceived Quality of Information in Wikipedia and Sharing Attitude.



3. Click OK to view output.

		Qu1	Qu2	Qu3	Qu5	SA1	SA2	SA3
Qu1	Pearson Correlation	1	.723**	.610**	.606**	.250**	.239**	.253**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000
	N	913	913	913	913	913	913	913
Qu2	Pearson Correlation	.723**	1	.595**	.536**	.250**	.223**	.227**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000
	N	913	913	913	913	913	913	913
Qu3	Pearson Correlation	.610**	.595**	1	.476**	.157**	.153**	.149**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000
	N	913	913	913	913	913	913	913
Qu5	Pearson Correlation	.606**	.536**	.476**	1	.275**	.251**	.257**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000
	N	913	913	913	913	913	913	913
SA1	Pearson Correlation	.250**	.250**	.157**	.275**	1	.611**	.635**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000
	N	913	913	913	913	913	913	913
SA2	Pearson Correlation	.239**	.223**	.153**	.251**	.611**	1	.616**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000
	N	913	913	913	913	913	913	913
SA3	Pearson Correlation	.253**	.227**	.149**	.257**	.635**	.616**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	913	913	913	913	913	913	913

Table 1: Correlation

In the output:

Pearson Correlation is the correlation coefficient between the items of interest.

Sig. (2 tailed) tells us whether the respective correlation is significant or not; and if significant, at what level of significance? Is it at p-value < 0.05 (designated by *), p-value < 0.01 (designated by **) or p-value < 0.001 (represented by ***).

N represents the sample size. In our case, it is 913.

We shall discuss this in detail in a future post.

You should by now realize the diagonal elements are the same. Reformatting the above table yields the table below.

Item	Perceived Quality of information in Wikipedia				Sharing Attitude		
	Qu1	Qu2	Qu3	Qu5	SA1	SA2	SA3
Qu1	1						
Qu2	.723**	1					
Qu3	.610**	.595**	1				
Qu5	.606**	.536**	.476**	1			
SA1	.250**	.250**	.157**	.275**	1		
SA2	.239**	.223**	.153**	.251**	.611**	1	
SA3	.253**	.227**	.149**	.257**	.635**	.616**	1

Table 2: Correlation

It is worth noting that for the majority of researches done in the social/behavioural sciences, discriminant validity is not tested at the item level as done above. Rather, researchers employ at least one of the following methods to confirm the distinct nature of their variables.

1. Cross loadings
2. Fronell-Larcker Criterion
3. Heterotrait-monotrait ratio of correlation (HTMT)

Confirmatory Factor Analysis (CFA)

Confirmatory Factor Analysis (CFA) is conducted after the researcher has gained substantial knowledge about the latent variables (s)he wishes to study. Thus, the researcher has already explored the underlying latent variables and know exactly the number of items/indicators that loads strongly on each of them via Exploratory Factor Analysis (EFA).

Unlike EFA, CFA aim at testing how well the data fits the proposed researched model. In other words, the researcher imposes the data on the research model and then determine the goodness-of-fit of this data to the model through the ‘eye’ of some measurement fit indices.

Convergent Validity

Convergent validity is a subset of construct validity. Here, the researcher's aim is to find out whether the items he claims are measuring a particular construct are indeed measuring them. In other words, if these items do measure a specific construct, then they need to converge.

The researcher achieves this by taking into consideration

	Rule of thumb
Factor loading	> 0.60
Cronbach's Alpha	> 0.70
Composite reliability	> 0.70
Average variance extracted	> 0.50

All these items have already been studied and we refer the reader to our post on that. Below is a summary of the table representing a good convergence fit for all the constructs. Thus, all the items are indeed measuring what they were intended to measure.

Construct	Item	Factor loading	Cronbach Alpha	CR	AVE
Quality (Qu)	Qu1	.893	0.851	0.883	0.658
	Qu2	.884			
	Qu3	.804			
	Qu5	.637			
Use Behavior (Use)	Use1	.798	0.862	0.880	0.648
	Use2	.857			
	Use3	.786			
	Use4	.776			
Incentive (Inc)	Inc1	.771	0.831	0.889	0.668
	Inc2	.814			
	Inc3	.892			
	Inc4	.787			
Sharing Attitude (SA)	SA1	.859	0.872	0.898	0.745
	SA2	.871			
	SA3	.859			
Perceived Ease of Use (PEU)	PEU1	.704	0.575	0.758	0.512
	PEU2	.681			
	PEU3	.759			

Factor loading and Cross-loading

Because constructs are measured indirectly, it is recommended that you have at least three items/indicators/questions measuring each of them.

For example, in the data used for this post, university faculty members were asked about their perceived quality of information in Wikipedia using the following five different questions.

QU1: Articles in Wikipedia are reliable

QU2: Articles in Wikipedia are updated

QU3: Articles in Wikipedia are comprehensive

QU4: In my area of expertise, Wikipedia has a lower quality than other educational resources

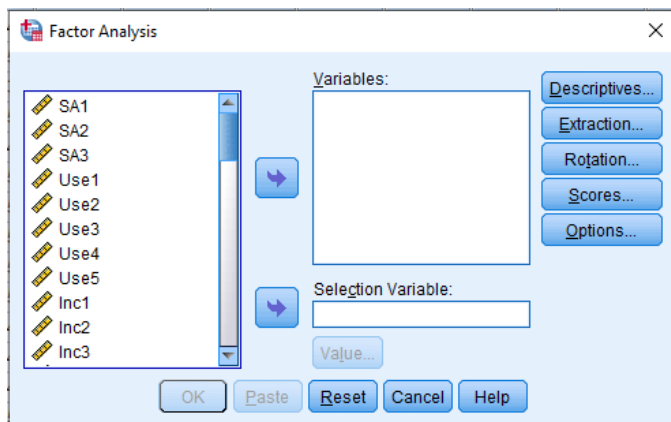
QU5: I trust in the editing system of Wikipedia

Their responses were recorded on a five-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5). So, all these questions are measuring the construct Quality and they are coded as QU1 (Question 1) to QU5 (Question 5). If these five items or questions are really measuring Quality,

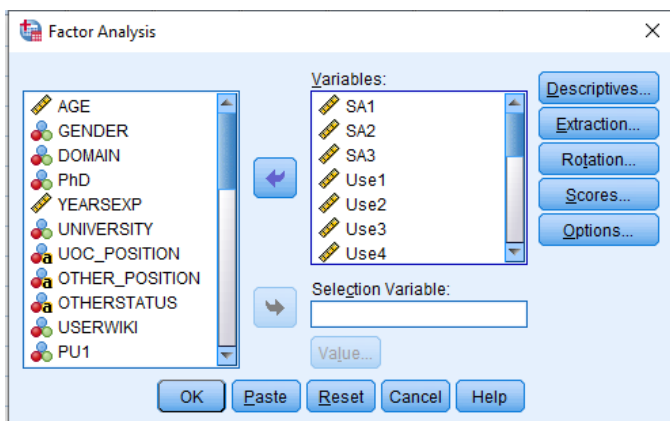
we must expect them to have a very high score on Quality. We call these scores, **factor loadings** or **loadings**. Preferably, we expect these loadings to be above the threshold of 0.6. Similarly, we shall expect these items to have very low loadings with other constructs, a term known as **cross-loadings**.

In this tutorial, we shall learn how to find the loadings and cross-loading for your data using SPSS. You can follow along by downloading the dataset.

1. In SPSS, click Analyze → Dimension Reduction → Factor. You will be presented with the window below.

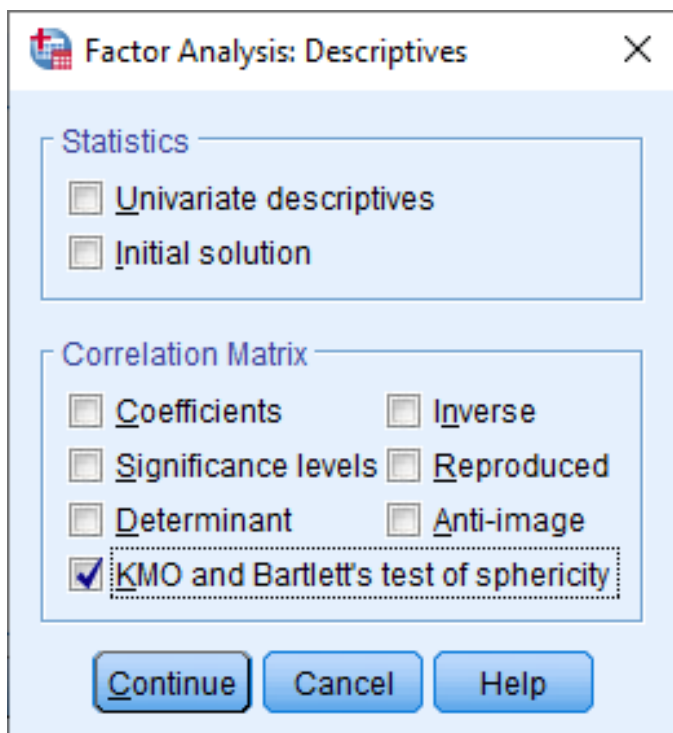


2. Move all the items measuring a particular construct into the Variables: box.

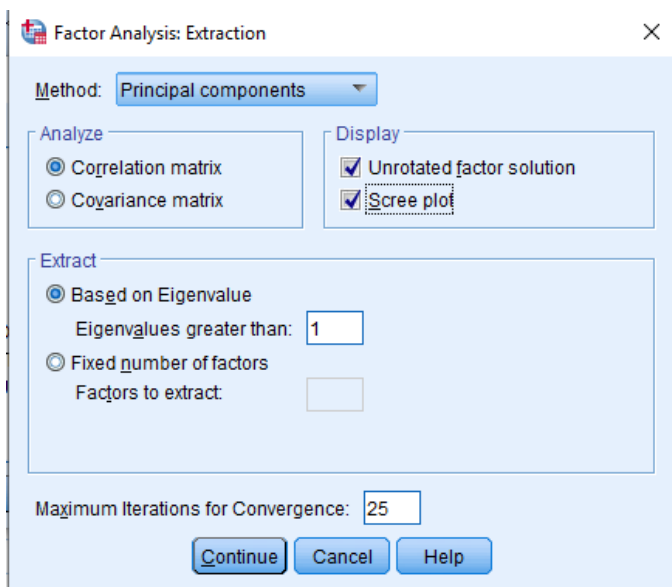


3. Click on;

(a) Descriptives: select **KMO and Battlett's test of Sphericity**, then click Continue



(b) Extraction: select **Scree Plot**, click Continue



(c) Rotation: select **Promax**, click Continue

Factor Analysis: Rotation

Method

☐ None
 ☐ Quartimax
☐ Varimax
 ☐ Equamax
☐ Direct Oblimin
 ☒ Promax

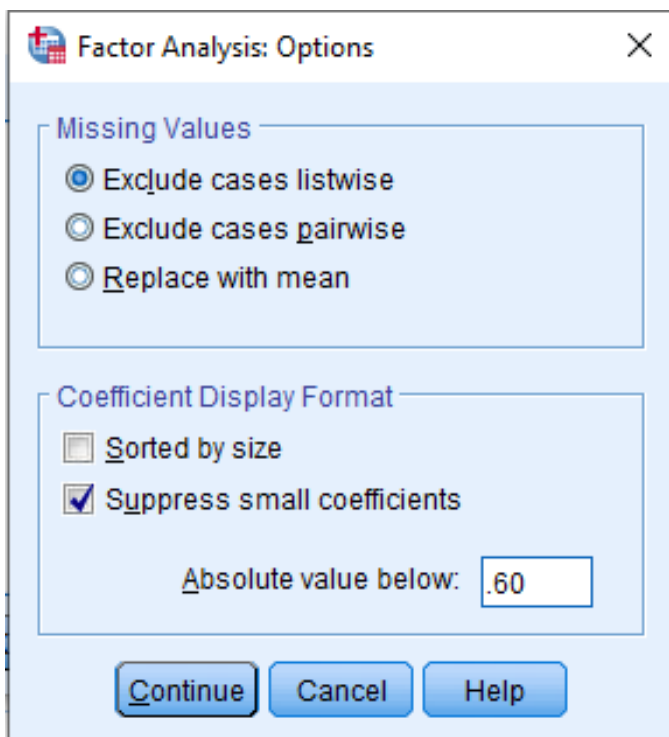
Delta:
 Kappa:

Display

☒ Rotated solution
 ☐ Loading plot(s)

Maximum iterations for Convergence:

(d) Options: This section is optional but very helpful at the initial stage of identifying high loadings; you can use it when you want to print out factor loadings that exceed a particular threshold. Tick **Suppress small coefficients**, then under ***absolute value below**: change 0.10 → 0.60 because we don't want to print out loadings below 0.6.



4. Click Continue and then OK.

You will be presented with so many tables. At the moment, our focus is on loadings and cross loadings so lookout for **Pattern Matrix** (this is our factor loadings and cross-loadings).

With the exception of the first column, the number of columns you have must be the same as the number of constructs. For example, in the questionnaire, we have five constructs (see data source for description), hence we must expect five different columns. Each column is representative of

each construct. The idea is to make sure all the items measuring a construct have higher loadings in ONLY one column (preferably greater than 0.6) and smaller loadings with other constructs. If an item or items from a particular construct measure strongly (above 0.60) on another construct, delete (remove) that (those) item (s) from your analysis in Step 2, one at a time. For each item removed rerun the analysis, repeating Step 2 until you get the desired output. The minimum number of items for each construct must be 3.

Pattern Matrix ^a					
	Component				
	1	2	3	4	5
SA1				.851	
SA2				.840	
SA3				.840	
Use1	.796				
Use2	.840				
Use3	.803				
Use4	.788				
Use5					
Inc1			.772		
Inc2			.810		
Inc3			.891		
Inc4			.785		
Qu1		.910			
Qu2		.901			
Qu3		.816			
Qu4					
Qu5		.640			
PEU1					.666
PEU2					.664
PEU3					.715

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.

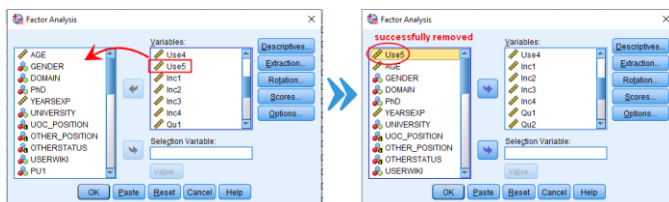
a. Rotation converged in 6 iterations.

From the Pattern Matrix table, you can see that

Use5: I agree my students use Wikipedia in my courses and

Qu4: In my area of expertise, Wikipedia has a lower quality than other educational resources

have no loadings (just that their loadings are below 0.6), so we delete them one at a time. You can do this by first removing either of them from the Variables section (see the figure below for details). Repeat this for Qu4.



The final output is shown below.

Pattern Matrix^a

	Component				
	1	2	3	4	5
SA1				.859	
SA2				.871	
SA3				.859	
Use1		.798			
Use2		.857			
Use3		.786			
Use4		.776			
Inc1			.771		
Inc2			.814		
Inc3			.892		
Inc4			.787		
Qu1	.893				
Qu2	.884				
Qu3	.804				
Qu5	.637				
PEU1					.704
PEU2					.681
PEU3					.759

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Great, now everything looks good.

Now, in step 3 (d), we can untick **Suppress small coefficients**, re-arrange the variables and then rerun the analysis to get the output below.

Pattern Matrix^a

	Component				
	1	2	3	4	5
Qu1	.893	.011	.002	.020	-.041
Qu2	.884	-.049	.013	-.010	.006
Qu3	.804	.032	-.017	-.149	.097
Qu5	.637	.218	-.032	.125	-.107
Use1	.069	.798	.000	-.038	.018
Use2	-.131	.857	.000	-.089	.030
Use3	.126	.786	.003	.099	.006
Use4	.125	.776	.012	.063	-.003
Inc1	-.058	.033	.771	.067	.094
Inc2	.078	-.028	.814	-.019	-.057
Inc3	.015	-.022	.892	-.057	-.048
Inc4	-.060	.032	.787	.033	.012
SA1	-.013	.005	-.002	.859	.024
SA2	-.032	.005	.005	.871	-.035
SA3	-.023	-.014	.009	.859	.040
PEU1	.083	-.209	-.041	.160	.704
PEU2	.295	-.082	.060	-.052	.681
PEU3	-.255	.303	-.019	-.062	.759

Extraction Method: Principal Component Analysis.

Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

The table above doesn't look very much presentable and quite confusing. Isn't it? Yes, so you can copy the output into any word document then format it to look beautiful and very much presentable as shown below.

Construct	Item	QU	Use	Inc	SA	PEU
Quality (Qu)	Qu1	.893	.011	.002	.020	-.041
	Qu2	.884	-.049	.013	-.010	.006
	Qu3	.804	.032	-.017	-.149	.097
	Qu5	.637	.218	-.032	.125	-.107
Use Behavior (Use)	Use1	.069	.798	.000	-.038	.018
	Use2	-.131	.857	.000	-.089	.030
	Use3	.126	.786	.003	.099	.006
	Use4	.125	.776	.012	.063	-.003
Incentive (Inc)	Inc1	-.058	.033	.771	.067	.094
	Inc2	.078	-.028	.814	-.019	-.057
	Inc3	.015	-.022	.892	-.057	-.048
	Inc4	-.060	.032	.787	.033	.012
Sharing Attitude (SA)	SA1	-.013	.005	-.002	.859	.024
	SA2	-.032	.005	.005	.871	-.035
	SA3	-.023	-.014	.009	.859	.040
Perceived Ease of Use (PEU)	PEU1	.083	-.209	-.041	.160	.704
	PEU2	.295	-.082	.060	-.052	.681
	PEU3	-.255	.303	-.019	-.062	.759

Table 1: Loadings and Cross Loadings

Under constructs we represented Qu1, Qu2, Qu3 and Qu5 as Quality (Qu) etc. The columns have also been labeled accordingly to reflect their exact meaning. Bolded numbers are the factor loadings, otherwise cross-loading

Table 1 gives an overview of the items that measure highly on a construct. Once this is done, you will be able to decide which question(s)/item(s) in your questionnaire do not measure what it was intended to measure. This could be due to misunderstanding from the respondent while filling the questionnaire or the question was not well structured by the researcher. If you are able to get this out of your

data, then you may proceed with your exploratory factor analysis (EFA).

How to calculate Average Variance Extracted and Composite Reliability

1. Select the number of available indicators (see figure below). For example, we have four items or indicators measuring perceived quality of information in Wikipedia (Qu1, Qu2, Qu3 and Qu5), so we selected 4 indicators as shown below.



2. Copy your factor loadings and paste them in the corresponding cell.

	Factor loading	SMC	1-SMC	CR	AVE
Item 1	0.893	0.797	0.203	0.883	0.658
Item 2	0.884	0.781	0.219		
Item 3	0.804	0.646	0.354		
Item 4	0.637	0.406	0.594		
					
Composite Reliability, CR			0.883		
Average Variance Extracted, AVE			0.658		
A composite reliability of 0.70 is recommended(Hair,1997) , Fornell and Larcker (1981) recommended a CR value of 0.60 or more Fornell and Larcker(1981) recommended an AVE greater than 0.5					

3. Record your AVE and CR
4. Repeat this process for all your constructs.

Note: Each construct must have an AVE and a CR

The table below shows the output for the AVE and CR of our study

Construct	Item	CR	AVE
Quality (Qu)	Qu1	0.883	0.658
	Qu2		
	Qu3		
	Qu5		
Use Behavior (Use)	Use1	0.880	0.648
	Use2		
	Use3		
	Use4		
Incentive (Inc)	Inc1	0.889	0.668
	Inc2		
	Inc3		
	Inc4		
Sharing Attitude (SA)	SA1	0.898	0.745
	SA2		
	SA3		
Perceived Ease of Use (PEU)	PEU1	0.758	0.512
	PEU2		
	PEU3		

Discriminant Validity through Cross Loadings

According to Gefen and Straub (2005), “discriminant validity is shown when each measurement item correlates weakly with another construct excepts for the ones to which it is theoretically associated”.

In cross-loadings, the researcher examines the various items to identify those that have high loadings on the same construct and those that load highly on multiple constructs. Thus, to establish discriminant validity at the item level means there is a high correlation between items of the same construct and a very weak correlation between items of a different construct. As simple as this approach is, it has no theoretical justifications or empirical proof (Henseler et. al., 2015).

We refer our reader here on how to find cross-loading. In the table below, the bolded items (in blue) represent the factor loadings for each construct and the cross-loading are those in red for the same construct. You can see the cross-loading

for each construct is very low indicating good discriminant validity.

Construct	Item	QU	Use	Inc	SA	PEU
Quality (Qu)	Qu1	.893	.011	.002	.020	-.041
	Qu2	.884	-.049	.013	-.010	.006
	Qu3	.804	.032	-.017	-.149	.097
	Qu5	.637	.218	-.032	.125	-.107
Use Behavior (Use)	Use1	.069	.798	.000	-.038	.018
	Use2	-.131	.857	.000	-.089	.030
	Use3	.126	.786	.003	.099	.006
	Use4	.125	.776	.012	.063	-.003
Incentive (Inc)	Inc1	-.058	.033	.771	.067	.094
	Inc2	.078	-.028	.814	-.019	-.057
	Inc3	.015	-.022	.892	-.057	-.048
	Inc4	-.060	.032	.787	.033	.012
Sharing Attitude (SA)	SA1	-.013	.005	-.002	.859	.024
	SA2	-.032	.005	.005	.871	-.035
	SA3	-.023	-.014	.009	.859	.040
Perceived Ease of Use (PEU)	PEU1	.083	-.209	-.041	.160	.704
	PEU2	.295	-.082	.060	-.052	.681
	PEU3	-.255	.303	-.019	-.062	.759

Loadings and Cross loadings

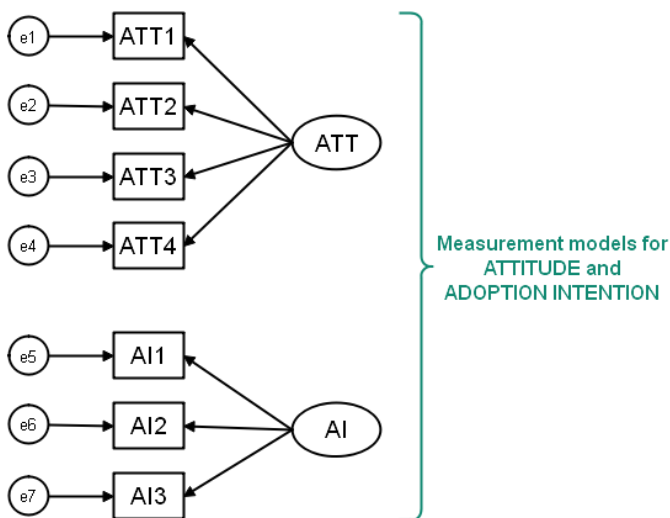
Measurement versus Structural Model

In Structural Equation Modeling (SEM), we can think of a model in two perspectives:

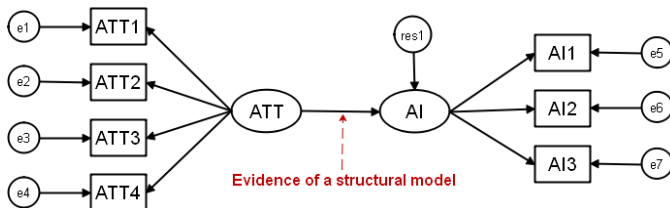
1. Measurement Model
2. Structural Model

The *measurement model* deals with the relationship between a latent variable and its indicators.

For example: In measuring the ATTITUDE(ATT) of employees towards the ADOPTION INTENTION of new technology at the workplace, the researcher has four (4) and three (3) items with good loadings (> 0.70) measuring respectively the latent variables ATTITUDE i.e. ATT1 - ATT4; and ADOPTION INTENTION i.e. AI1-AI3. So, the link between each of these items/indicators with their respective latent variable is known as a measurement model.



In contrast, a structural model defines the relationship between the various constructs in a model. In the example above, the two measurement model becomes a structural model when they are linked together as shown below. Thus, specifying how latent variables directly or indirectly affect other latent variables in the model.



Average Variance Extracted (AVE)

When a researcher finds the average variance extracted for a construct, he is interested in knowing, on average, how much variations in his items can be explained by the construct or latent variable.

For instance: in the table below, perceived quality of information in Wikipedia (Qu) was measured with four items and the AVE for these items is 0.658, which means, on average, 65.8% of the variations in university faculty's perceived quality of information in Wikipedia is explained by these four items or questions. On the other hand, we recall a 34.2% error in our measurement items which is fairly okay.

Construct	Average Variance Extracted
Quality (Qu)	0.658
Use Behavior (Use)	0.648
Incentive (Inc)	0.668
Sharing Attitude (SA)	0.745
Perceived Ease of Use (PEU)	0.512

As a rule of thumb and for adequate convergent, an AVE of at least **0.50** is highly recommended. That been said, an AVE less than 0.50 means your items explain more errors than the variance in your constructs. For any measurement model, an AVE must be calculated for each construct and must be at least 0.50.

Composite Reliability

This is another way a researcher can measure the internal consistency of his items. It is recommended that the reliability of a construct is at least 0.70. High composite reliability is a very good indication that all your items constantly measure the same construct. From the results below, we could see all the composite reliability for the constructs ranges from 0.758 to 0.898, which exceeds 0.70. A clear indication that all the items consistently measure their corresponding construct.

Construct	Composite Reliability
Quality (Qu)	0.883
Use Behavior (Use)	0.880
Incentive (Inc)	0.889
Sharing Attitude (SA)	0.898
Perceived Ease of Use (PEU)	0.758

Latent versus Observed variables

Consider the following questions asked by a researcher in a questionnaire on Social Acceptance:

1. The product should be approved by my friends.
2. The product should be popular in my circle of friends.
3. The product should make a good impression on people who are important to me.
4. The product should be liked by people who are important to me.
5. The product should be liked by people who are important to me.

I am pretty sure you have come across several of these examples in any survey questionnaire. In this example, Social Acceptance is measure by five items.

Why?

This is because Social Acceptance by itself is too broad and the researcher cannot measure it directly, so he or she is tasked to design a set of questions/items (that have been empirically tested and conforms to theory) to measure Social Acceptance. These sets of questions/items are called observed/manifest variables. In Structural Equation Modeling (SEM), they are usually called indicators. Social Acceptance, on the other hand, is known as a construct/factor/latent/unobserved variable. Other latent variables include Quality, Hedonic Motivation, Pleasure, Attitude. and so on.

Exploratory Factor Analysis (EFA)

In Exploratory Factor Analysis (EFA), the researcher has no prior intention of confirming the existence of any existing relationship prior to the analysis of the data, but instead, the researcher allows the various techniques, methods, procedures and rules to reveal the relationship in the data. EFA is indeed the opposite of Confirmatory Factor Analysis (CFA).

In the Social/Behaviour Sciences, the study of EFA comprises of the following, the list, however is not exhaustive.

- KMO and Bartlett's test of sphericity

- Communalities

- Scree Plot

- Loadings and Cross loadings

- Reliability Analysis

- Composite Reliability

- Average Variance Extracted (AVE)

- Convergent Validity

- Discriminant Validity

End of the Book