Marketing Research and Analysis -II (Application Oriented)
Prof. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of Technology - Roorkee

Lecture – 54
Confirmatory Factor Analysis in SPSS - I

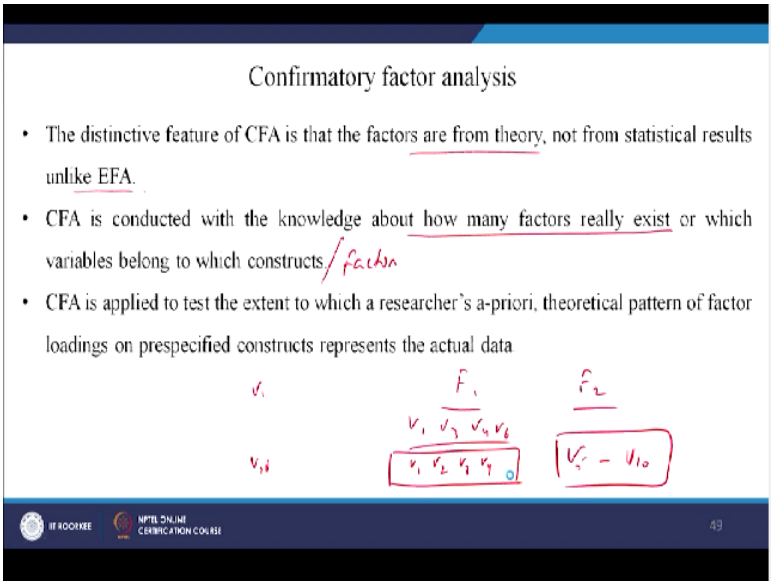
Welcome everyone to the class of marketing research and analysis. In the last lecture, we had discussed about factor analysis. Especially we talked about the exploratory factor analysis in which we learnt how to statistically derive a few factors out of a large number variables, right. And how it is used also we learnt that those factors can be later on used in some other studies and other techniques for example in regression or in anywhere and can be used for that purpose, okay.

But the way we conducted that study was exploratory in nature. Because we had no clue. We had no idea, right, what exactly we were doing. So we were finding the factors on the basis of certain measures like for example, the variance, the eigenvalue, we are extracting on basis of, for example, how much of variance is being explained in the study. So there were certain measures on which we were selecting the factors, right.

So we were looking into the correlation among the variables and trying to see that they were of the same factor or not. But today, we will be talking about another technique under factor analysis which is called confirmatory factor analysis. And why it is called confirmatory, let us understand. So earlier you realized that it was called exploratory because you were exploring, right. But here you are not exploring anything.

You are confirming. So what do you confirm? You can only confirm something when there is prior knowledge, right. So when you have prior knowledge, you are only testing it to check whether it is coming true or not. So in such conditions, we say it is being confirmed or it is a confirmatory test, okay. So similarly, this technique is called confirmatory factor analysis. And what it is, let us see?

(Refer Slide Time: 02:16)

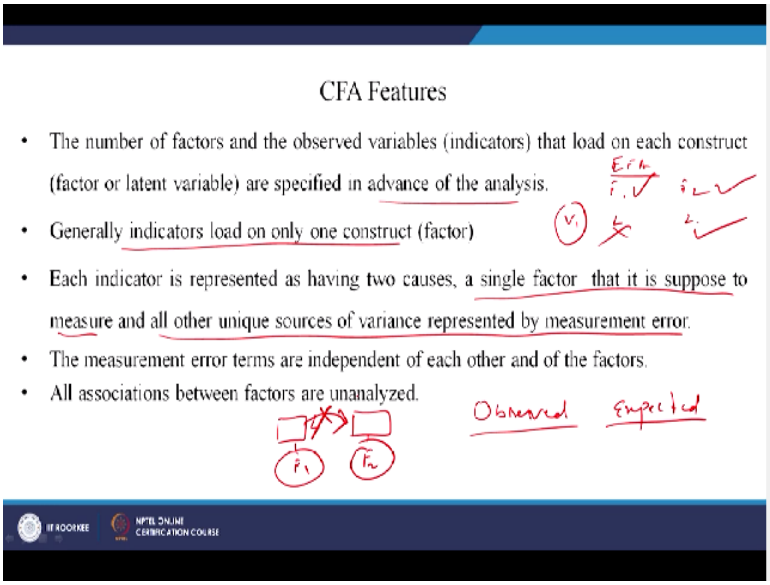


The distinctive feature of CFA is that the factors are from theory. So not from statistical results unlike EFA. So that means the researcher has to have a prior knowledge whether how these variables should be grouped into a factor. So this knowledge, they should have it from the theoretical background, from the literature review or something, okay. CFA is conducted with the knowledge about how many factors really exist or which variables belong to which construct or factor.

This word construct or factor are interchangeably used, okay. CFA is applied to test the extent to which a researcher's a-priori, that means beforehand, what knowledge he had, theoretical pattern of factor loadings on the pre-specified constructs represents the actual data. That means what it is being said is suppose there are two factors. So in the explanatory factor analysis, we had, let us say, V1 to V10.

And we, after doing the study, we found that, let us say, V1, V3, V4, V6 were loaded into factor 1 and the remaining were into factor, let us say, 2. So this was from the factor loadings, we found out the variables connection with the factors. But during the confirmatory factor analysis, we have two pre-specified. For example, we will say that V1, V2, V3 and V4 are into factor 1. So this has to be pre-determined. Already you have to say beforehand from the theoretical perspective. And V5 to V10 will lie in the factor 2, okay.

(Refer Slide Time: 03:58)



The number of factors and the observed variables that load on each construct are specified in advance of the analysis, right. So you have no scope to do anything after the test is done. So you have to, that means what the researcher has to say, how the model will be beforehand so that what he is thinking today or now and after the model has been statistically checked, both the results should come true. That means what?

Suppose there are two things here. One is called the observed, right and the other is called the expected. As you must have heard this term while doing chi square also. So same thing happens here. So we have an observed model, we have an expected model or a predicted model, right. So these two models, right, so we have to have a very close difference. That means that difference between the observed and the expected has to be very low.

Then only we can say that we expected and what we observed is the same. That means we are confirming that it is true, right. That is the meaning, okay. So generally the indictors load on only one construct. And if you remember in exploratory factor analysis, we had seen that sometimes, suppose there is a factor 1 and factor 2. Suppose the variables, sometimes the V1 was loaded here also and here also.

It was in both the factors. So in that condition, we said that either it is an item for deletion, that means if you have sufficient number of variables, it is better to delete that particular variable

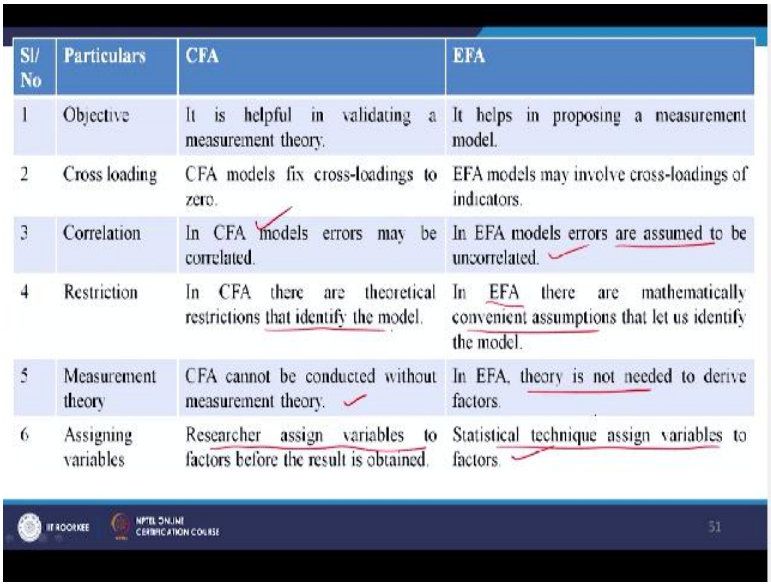
from the study. Or if you have small number of variables, then what you should do is? You should understand that theoretically whether V1 should fall in F1 or F2, theoretically, logically. So if it falls in F2, then it should be taken to F2 and this should be ignored, right.

So this is what we did in EFA. But that condition completely is now out of question in the CFA. Because in the CFA, already you have said I have a theoretical understanding from the literature or from some experience or something, right. So here, the indicators load only on one construct. So no question of cross loading, okay. Each indicator is represented as having two causes. A single factor that it is supposed to measure and all other unique sources of variance represented by the measurement error.

So there is one thing called the measurement error which we show it like a rectangle. And then there is a construct, right, suppose to which is, it is suppose to measure, the construct, right. So these are the two things that we will be checking here. The measurement error terms are independent of each other. So let us say, there is a construct. So each construct has a measurement error term.

Let us say this is F1, F2, right. So each has a measurement error term. And they are independent of each other, right. So they are not connected. There is no connection between them, right. There is no connection, right. All associations between factors are analyzed. So all the associations between the factors are analyzed. That means between the variables, the different items, it is checked.

(Refer Slide Time: 07:14)



Sl/ No	Particulars	CFA	EFA
1	Objective	It is helpful in validating a measurement theory.	It helps in proposing a measurement model.
2	Cross loading	CFA models fix cross-loadings to zero.	EFA models may involve cross-loadings of indicators.
3	Correlation	In CFA models errors may be correlated.	In EFA models errors are assumed to be uncorrelated.
4	Restriction	In CFA there are theoretical restrictions that identify the model.	In EFA there are mathematically convenient assumptions that let us identify the model.
5	Measurement theory	CFA cannot be conducted without measurement theory.	In EFA, theory is not needed to derive factors.
6	Assigning variables	Researcher assign variables to factors before the result is obtained.	Statistical technique assign variables to factors.

Let us see the difference between CFA and EGA, objective. It is helpful in validating a measurement theory, correct. You have a measurement, you have some theory in mind. You want to validate whether it is true or not. So that is why you create a model beforehand. Here in EFA, exploratory, it helps in proposing a measurement model. So that is why the EFA comes before the CFA.

So it helps in proposing. After you have done a study, statistical analysis, you said that factor 1 consists of these variables, factor 2 consist of these variables and now the relationship between this is something like this, okay. Cross loading. CFA models fix cross loadings to 0. So there is no cross loading as such. But EFA models may involve cross loadings of indicators. We have seen that.

In CFA models, errors may be correlated. So if there are errors, so there will be some error term. What is the error term? In the earlier classes, we have tried to explain error terms are nothing but the unexplained variance. So this unexplained variance, there might be some correlation as per the CFA models. But in the EFA, errors are assumed. In practicality, it is not possible. But they are assumed to be.

That means the entire spirit of developing the EFA, they assume that the errors are uncorrelated, okay. In CFA, there are theoretical restrictions that identify the model. So there is a theoretical

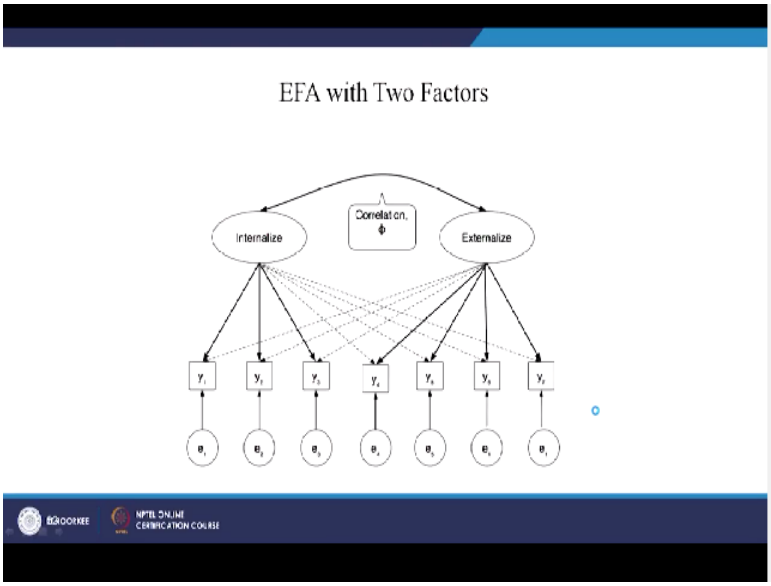
restriction, okay. In the EFA, it is a mathematically convenient assumption. So mathematically convenient, we say all the factors, all the variables are loading into this factor, there is a convergent variety.

There is a high loading, okay. CFA cannot be conducted without measurement theory, we have understood that. In EFA, theory is not needed to derive factors. It is through the loadings or the, how much of loading it is on the factor, right. Last, the researcher assign variables to factors before the result is obtained. Now this is very important, you see. the researcher is assigning the variables to the factors before the result.

So that means may be he has seen some other researcher doing a similar study and that is, he has already, let us say that work is published or it is a validated work, then that particular construct or scale, now I am suppose using. Then what is happening here? The variables to the factors are already known to me. Here, strategical technique, assign variables to the factors. So we have done it through the different ways through eigenvalues, through extraction factors.

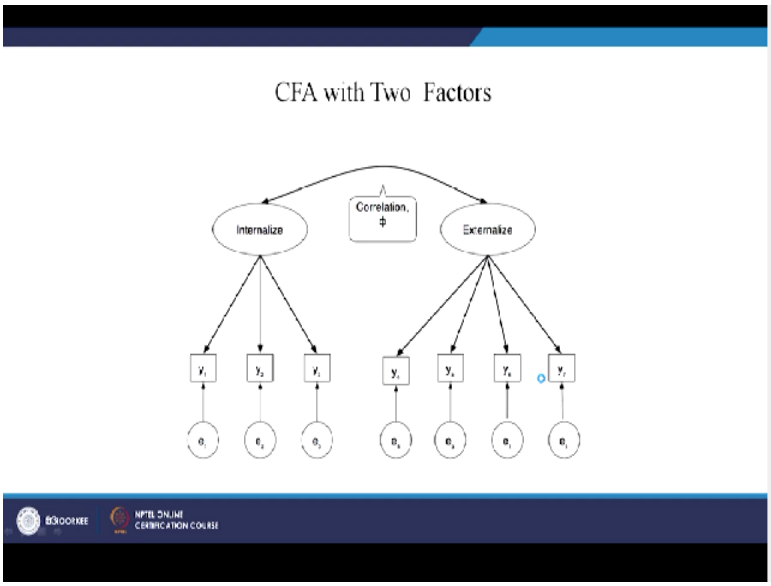
So we have understood how to assign the variables to the factors, okay. So this is the major difference. So just if I tell you the nutshell, the major difference is EFA comes through statistically and you are exploring here. CFA, it comes from theory and the researcher has to make the theory before he goes for the statistical interpretation.

(Refer Slide Time: 10:12)



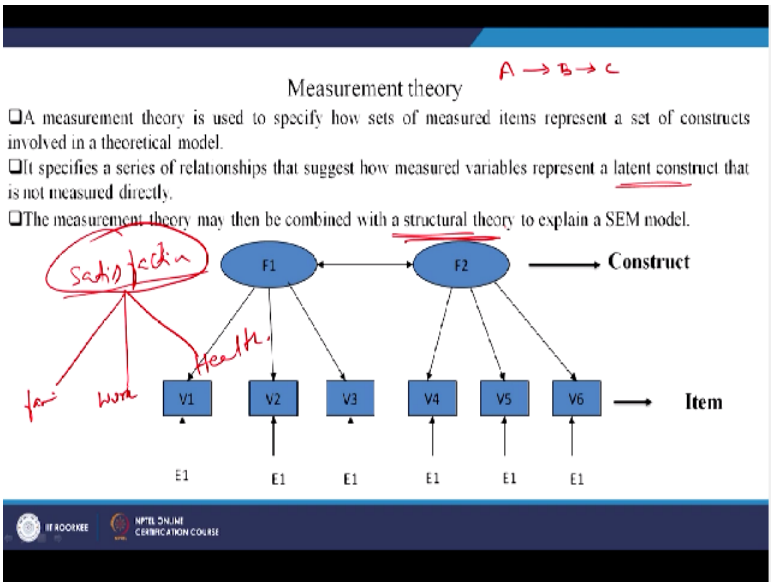
Now let us see this EFA with two factors. How does the EFA look? So there are two constructs. Construct 1, construct 2, okay. Now and these are the different, the variables or items, whatever you say, right. Now look at here. There is a chance of cross loading. Can you see? So from construct 1, there may be some variables loading into this factors also, this factor also, right. So there might be a cross loading case, right. So this is an EFA.

(Refer Slide Time: 10:41)



Same thing now coming to the CFA. So there are two constructs and you see the variables are clearly assigned to one and only one construct. And there is no chance of cross loading. So this is one of the major differences, right. And these are by measurement, the error terms, right, measurement errors, okay.

(Refer Slide Time: 11:00)



What is a measurement theory? A measurement theory is used to specify how sets of measured items represent a set of constructs involved in a theoretical model. In this case, there are two constructs, right, F1 and F2. And this is a, remember CFA is a covariance model. Now what is a covariance? The difference between a covariance and correlation, I think is understood by now. In the earlier classes also, we have discussed many a times.

A covariance, a correlation is nothing but a standardized covariance, right. So a covariance means the two variables change their values when there is change in one and the other, right. So that means when both are changing according to their movement. When there is a movement in one, there is also a movement in the other. So we say it is a covariance structure, okay. Now look at these variables.

And there are the error terms, okay. So these are the items or the variables. It specifies a series of relationships that suggest how measured variables represent a latent construct. Now what is this term latent construct? Very important to understand. A latent construct is the term which we use while doing a CFA or SEM. Now what is latent? Latent means hidden. Sometimes or many a times in psychology, we cannot directly measure a variable or a construct, right.

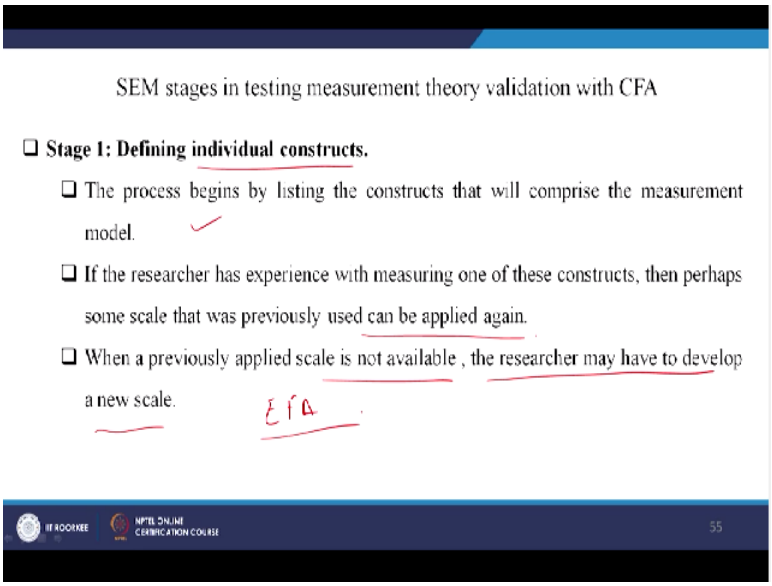
So to measure that, we have to ask several other questions which are related to that major

construct. Let us say suppose F1 is a latent construct. And F1 is let us say satisfaction, right. So to measure satisfaction, I cannot measure it directly. So I am measuring through 3 different items, V1, V2, V3. So the satisfaction could be family, right. Could be a work place, right. Or could be, let us say, my health, right.

So these are the 3 indicators I have used for measuring satisfaction. And 3 indicators together will explain the satisfaction. That is why satisfaction is called my latent construct, right. The measurement theory may then be combined with a structural theory to explain a SEM model. Now what is this? This is called a measurement model. But when we will try to use it, this is called a structural theory.

That means one when we establish relationships, so let us say, A affects B, B affects C. So when we establish such a relationship, then it becomes a structural model, right. So at this moment, we are not going to give any establish any relationship as such, okay. So this is called a measurement model. So CFA is a measurement model, remember.

(Refer Slide Time: 13:41)



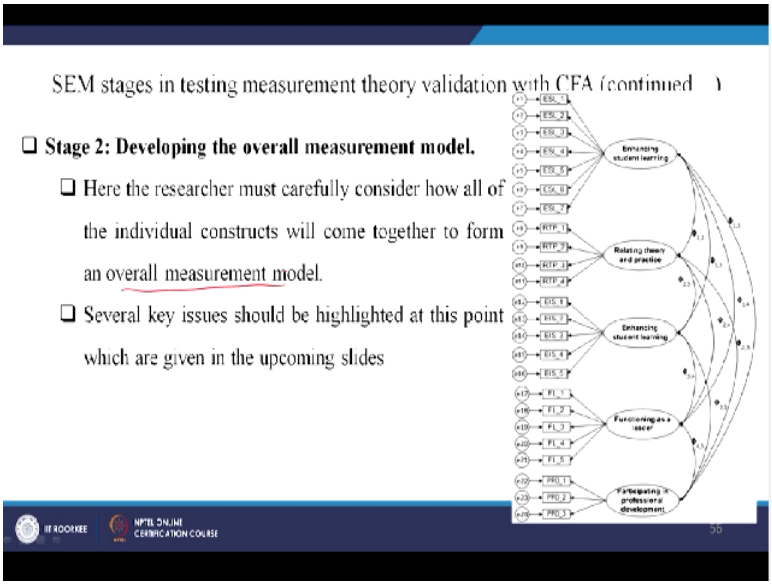
Now let us see the stages in testing measurement theory. So the first stage is to define the individual constructs. The process begins by listing the constructs that will comprise the measurement model. So in your model as then when you are doing a research, what are the different constructs that you would like to use, right. So you have to first keep it in mind, you

have to list it down, right.

And this will be completely in basis of your understanding, on the basis of your knowledge that you have gathered, okay. If the researcher has experienced that measuring one of this constructs, then perhaps some scale that was previously used can be applied again. Now this is what I was saying. Suppose somebody has a similar work in, let us say, Canada. But the thing is that scale is already validated.

So now I can use it in India also and see that whether the scale is working in the same manner or not, okay. When a previously applied scale is not available, is not available. So researcher sometimes ask this question, what will you do when the scale is not, my work is completely new. The researcher may have to develop a new scale. How does he develop a new scale? He can develop a new scale through the help of exploratory factor analysis. And then we can check for validity, okay. So this is a method you can adopt.

(Refer Slide Time: 15:05)

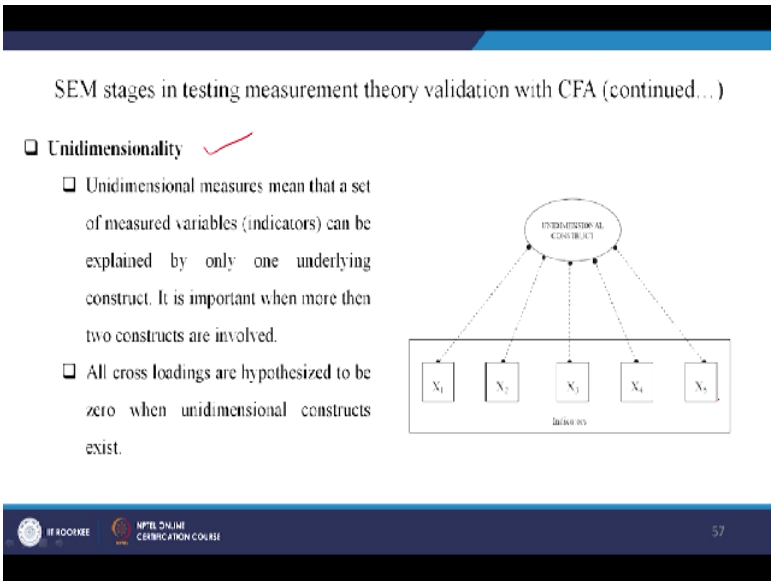


Stage 2. You develop the overall measurement model. Now you see, this is just an example. So you can see there are constructs. So construct 1, construct 2, construct 3, construct 4, construct 5. Each construct has got several variables. Can you see? So this one has got 1, 2, 3, 4, 5, 6, 7. So this one has got 1 to 4.

This one has got 5. So now after doing that and placing all the errors, right, related to each of the variables, then what we have done is. We have tried to build a covariance structure, now relating every construct with the other, right. Every construct has been connected to the other, right. So here the researcher must carefully consider how all of the individual constructs will come together to form an overall measurement model.

So here we are not establishing any relationship as such. We are trying to bring in all into one and trying to see that whether the entire model is a good model or not. To do that, we are connecting all the constructs with each other and trying to study it overall. Several key issues should be highlighted at this point which will be in the next slide.

(Refer Slide Time: 16:19)

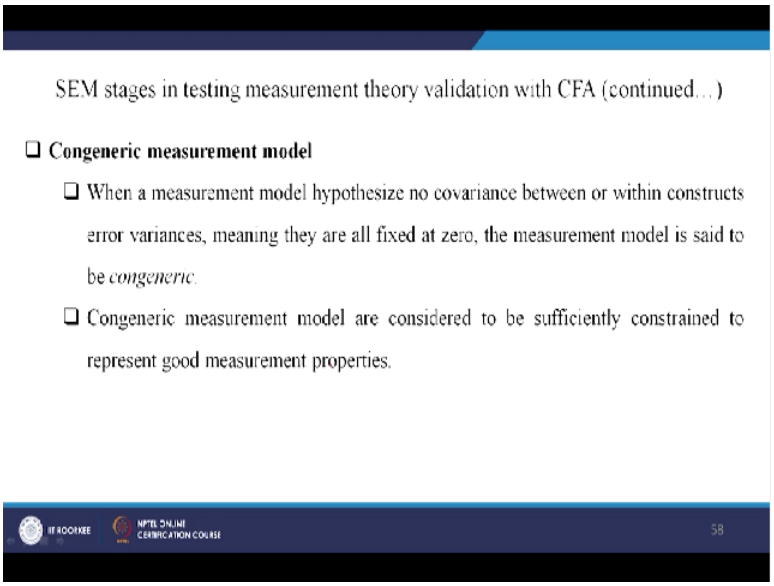


So let us see this. First is the case of unidimensionality. Now what does this mean? Unidimensional measures mean that a set of measured variables can be explained by only one underlying construct. It is important when more than two constructs are involved. This is the unidimensional construct, means what? Only one. these variables are loading into only one construct, right.

So that is why there is no scope of, this getting into some other constructs at all, right. All cross loadings are hypothesized to be 0. So the cross loading part is absolutely assumed to be 0. There is no cross loading when unidimensional construct exist, okay. So only one factor. So they are all

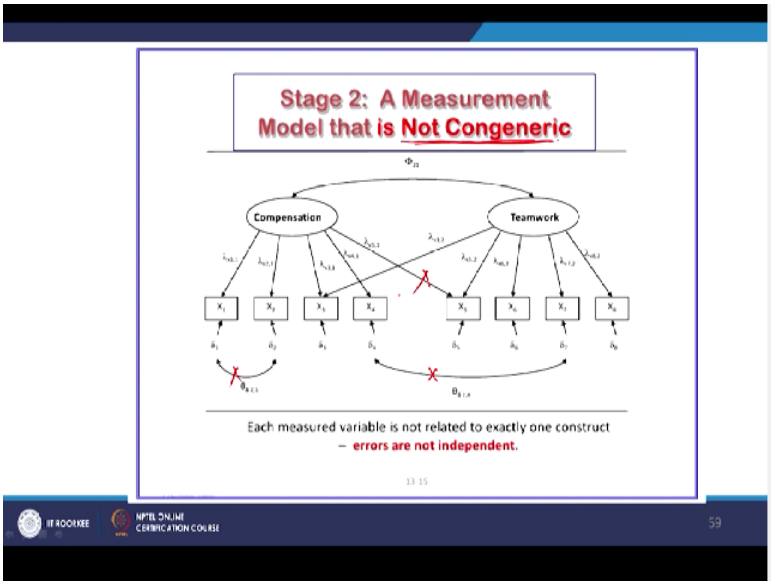
loaded, connected with only one factor. No chance of getting connected to some other factor, that is theoretically it is assumed to be not existing.

(Refer Slide Time: 17:13)



Second, congeneric measurement model. So what it is saying, when a measurement model hypothesize no covariance between or within the constructs error variances, meaning that they are all fixed at 0, the measurement model is said to be congeneric. Let us understand what it is. Congeneric measurement model are considered to be sufficiently constrained to represent good measurement properties.

(Refer Slide Time: 17:36)

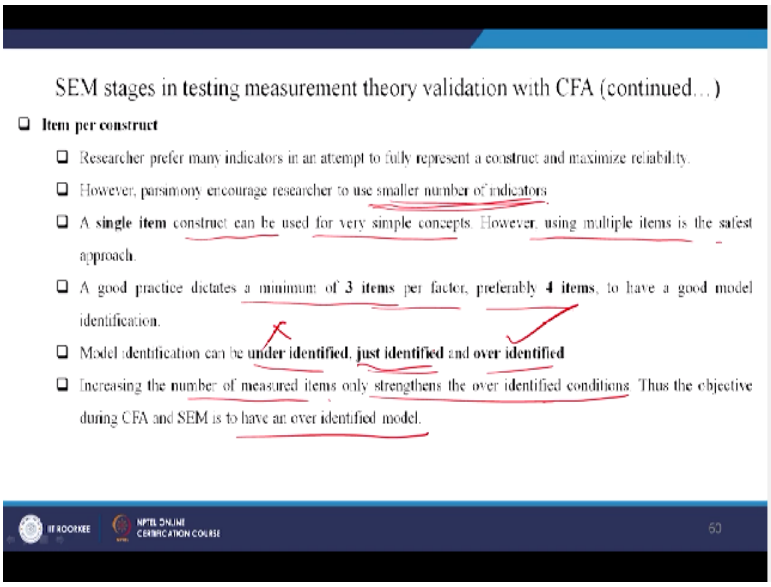


Now I think I have shown you, yes. Now you see this. Now each measured variable is not related

to exactly one construct, right, you see. So this is a not congeneric, remember it. If had been congeneric, that is a separate thing. We are taking, saying it is not congeneric. Why it is not congeneric, you see. First of all, this variable is loaded not only into these 4 but also has been loaded into a second factor here.

Second, you see among the error terms, there is a covariance, right. So here and here. So that is why as you go back, when a measurement model hypothesize no covariance but we have covariance. So it says that no covariance between or within construct should be there. But since it is there, we are saying it is not congeneric. Had this been not there, had this been not there and this been not there, then we could have said it is a congeneric model, okay.

(Refer Slide Time: 18:33)



Item per construct. Now many a times, this question comes to me that students ask sir how many items should be used. Researcher prefer many indicators in an attempt to fully represent a construct and maximize the reliability. So higher the number of items, right, or indicators, it is better. However, parsimony encourage researcher to use smaller number of indicators. See you have studied, I have told you earlier also, the research is parsimonious in nature.

What does it mean? You should be trying to always attain the highest amount of information with the lowest number of variables. So that is what researchers should try to achieve, okay. So if you use all variables that is possible, then it is like using a population or a census. Then what is

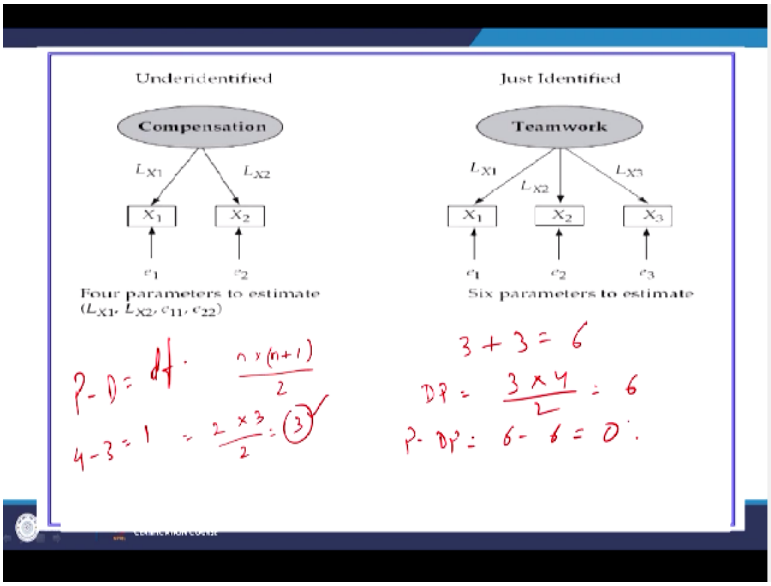
the point of conducting research? So we are trying to always see if that how does a sample behave or does the sample represent the population or not?

Similarly instead of using large number of indicators, can I get a good explanation by using a smaller number of indicators? so the smaller number of indicators must be the ones which actually do influence the measurement. A single item construct can be used for very simple concepts. See if the concept is very simple, then a single item is okay, fine. However, using multiple items is the safest approach.

Whenever you draw a construct or make a construct, you should have, it is better to have multiple items. That means 2 variables, 3 variables or 2 questions, 3 questions, whatever way you are attempting, right. A good practice dictates a minimum of 3 items per factor, preferably 4 to have a good model identification. Sometimes model can be under identified, just identified and over identified.

Now what does it mean? People must have heard, they must have read also, but sometimes they are very confused. So what does it mean, let me explain it. So increasing the number of measured items, when you increase the number of measured items, it only strengthens the over identified conditions. Thus the objective during CFA and SEM is to have an over identified model. So I will explain to you, I will show you that it is always better to have an over identified model than an under identified model, right. Just identified, it is also good, right.

(Refer Slide Time: 20:56)

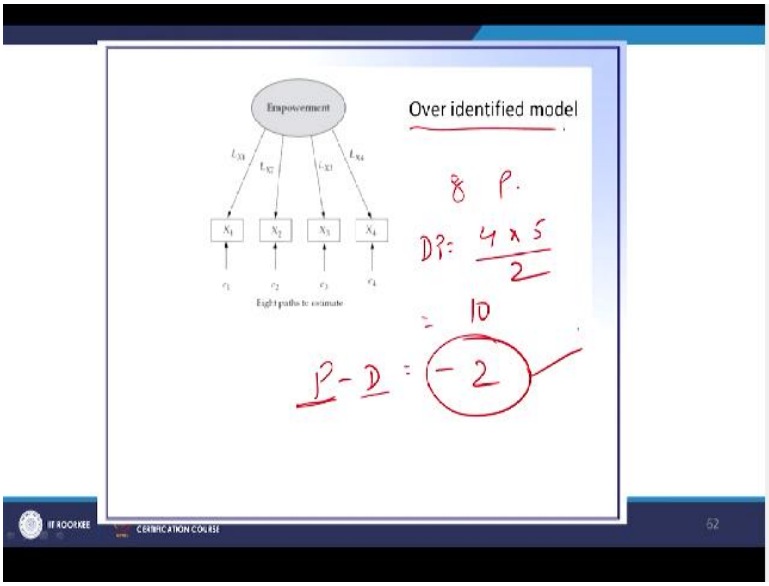


So this is the case of an under identified model. Now you see, how many, there are four parameters to estimate. 1, 2, 3, 4, right. So if you see there are four parameters to calculate, to estimate and in this, how many, if you calculate the number of parameters as per relationship, we calculate by a method called n*(n+1)/2. So in this case for example, how many variables are there? So 2. So 2*3/2, so is equal to 3.

So in this case, you have four parameters and you have got, as per calculation, you should have three. So in this case, what we are saying is, that if you look at the number of parameters and you calculate the number of data points, so which is actually the number of the covariances, right. So this if you divide this parameters minus the data points, so D is equal to, it gives us our degree of freedom. Now if this degree of freedom in this case for example you see 4 minus, it is 3, so=1.

So this is a case where we are saying it is an under identified modeling, okay. Similarly in this case, if there are, how many variables? 3+3=6. Now data points is equal to how much? 3*4/2, so how much? 6. So when, if you subtract this, parameters minus DP=6 - 6=0. So this is a just identified model, right.

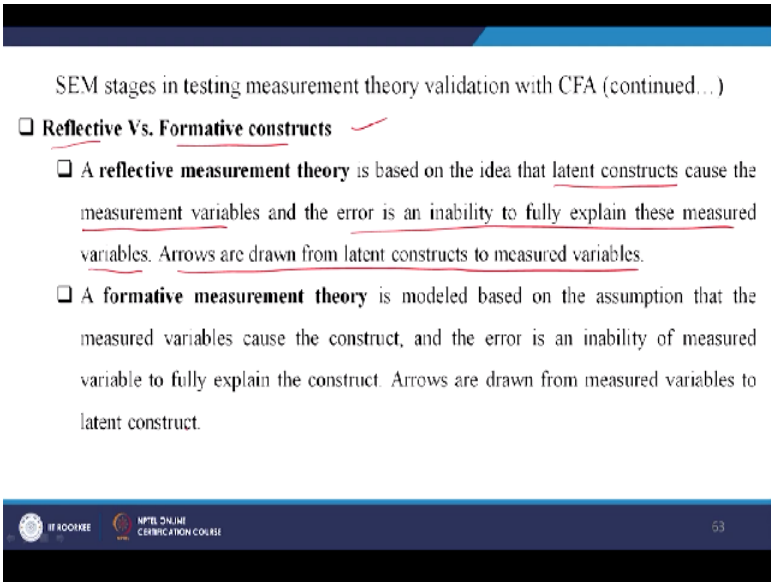
(Refer Slide Time: 22:42)



Now in the third case, this is an over identified model. Now let us look at it. Now 1, 2, 3, 4, 5, 6, 7, 8. So you have eight parameters, okay. Now the DP, data points is equal to how many? 4*5/2=how many? 10. So here, you have a P minus DP= -2, right. So this is a case of an over identified model. When the degrees of freedom, sorry, when the degree of freedom is negative, right or if you take it the other way round, it would be positive, so if you subtract it from parameters-data points, if it is negative, then it is called an over identified model, okay.

So if you see, whenever the number of variables is more than 4 or 4 even, it is, generally it is getting an over identified model, okay. So and you should always remember as a researcher that is why it is always said that the number of items per construct should be at least three or more. Otherwise, there is a chance that you find an under identified model and which will give you an improper solution, right. So that is very important to measure. Now coming to types of construct.

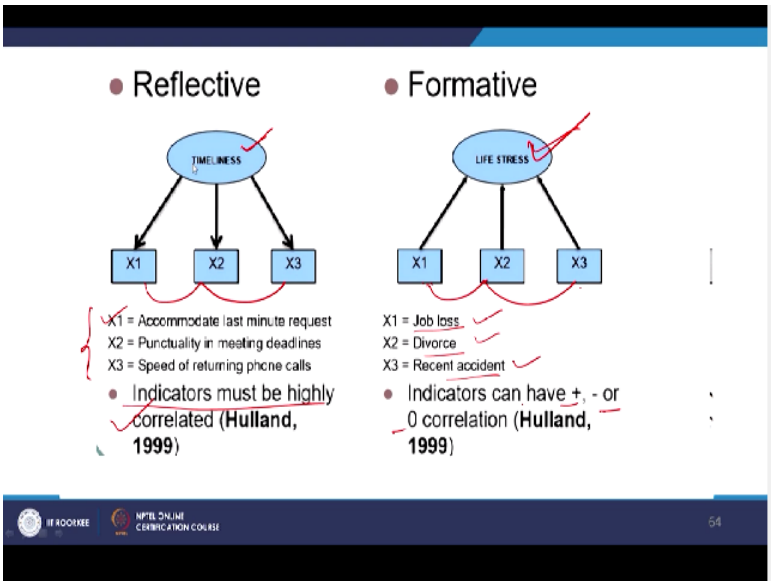
(Refer Slide Time: 23:56)



So there are two types of constructs. This is also a very important to understand. Reflective versus formative constructs. Now what are they? A reflective measurement theory is based on the idea that latent constructs cause the measurement variables and the error is an inability to fully explain these measured variables. Simple if you want to understand. How do you understand it that the arrows are drawn from the construct to the measured variables.

That means the construct is causing the variable, not the variable is causing the construct, right. I will show you an example also in the next slide. On the other hand, a formative theory is modeled based on the assumption that the measured variables cause the construct. This is generally what we understand, right. And the error is an inability of measured variable to fully explain the construct. Arrows are drawn from measured variables to the latent construct. Let us see the slide.

(Refer Slide Time: 24:53)



Reflective. Now we said that the construct explains the variables, right. And in formative, the variables explain the construct. Now let us take this example and understand. So timeliness, somebody wants to measure timeliness. So the 3 variables connected are, X1, X2, X3 accommodate the last minute request, punctuality in meeting deadlines, speed of returning phone calls.

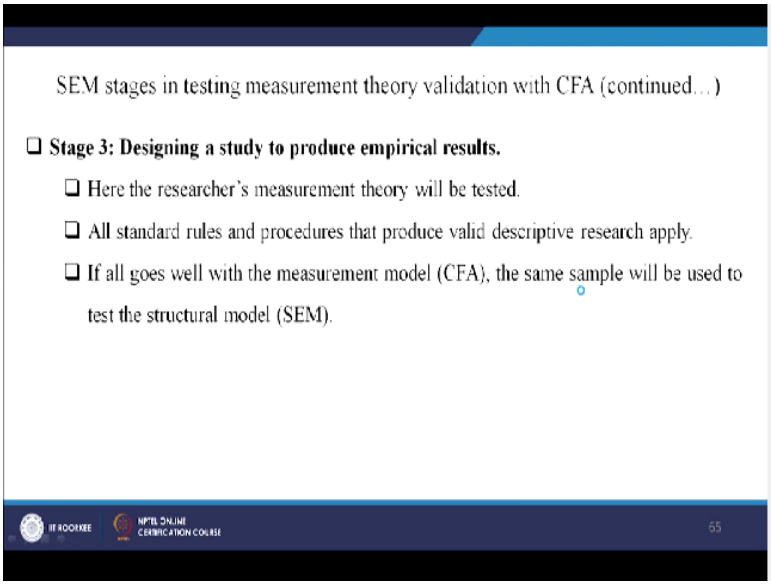
But if you logically think, these three are being explained through this construct timeliness, right. if somebody is timeliness, is following timeliness, then accommodating in the last minute request is possible. Punctuality in meeting deadlines is there. And speed of returning the phone calls is also there. That is the measure of timeliness, right. And here, you see the indicators must be highly correlated.

So these indicators must be highly correlated. But when it is a formative construct, what is happening, you see. Now life stress is being explained by these three. Here it was opposite. Timeliness was explaining these 3 or these 3 were being explained by timeliness. But here, these three are explaining life stress. Now what are they? For example, X1 is, this X1 now we are talking about.

Job loss, divorce, recent accident. So stress of a person, the life stress of a person is measured or is affected by job loss, divorce and recent accident. But it is not the other way round, right. That

means if life stress is there, that does not result in job loss and divorce and recent accidents. But we are explaining job loss, divorce and recent accidents in the way that these three together are explaining in the life stress. Indicators can have a positive, negative, or a 0 correlation. So these indicators might be positive related, might be negative related, or they might not be correlated at all. So these are the two types of constructs how you build, okay. Well, I will stop here.

(Refer Slide Time: 27:00)



We will continue in the next class where I will explain you how to carry on the study, right, how to carry on the study. And how to create a safer model out of your theory and how to interpret it and then from there, how do you establish relationships or how to develop a structural model. And explain the relationships between the different variables, okay. So this will be done in the next class, right. And we will also see by taking a data set an example, right where I will be drawing and I will be explaining you the details, okay. Thank you very much for the day.