Dr. Jogendra Kumar Nayak
Department of Management Students
Indian Institute of Technology Roorkee

Welcome friends to the class of marketing research and analysis. In the last session we disused about scale and what are the types of scale. Then what is the role of scale in the research, we have discussed about the importance of scale because scale is one such thing that helps in deciding what the type of statistical tool is? That one would use, to derive at the inference or to do some calculation or something right. So the type off test basically used statistical test depends on.

(Refer Slide Time: 01:00)
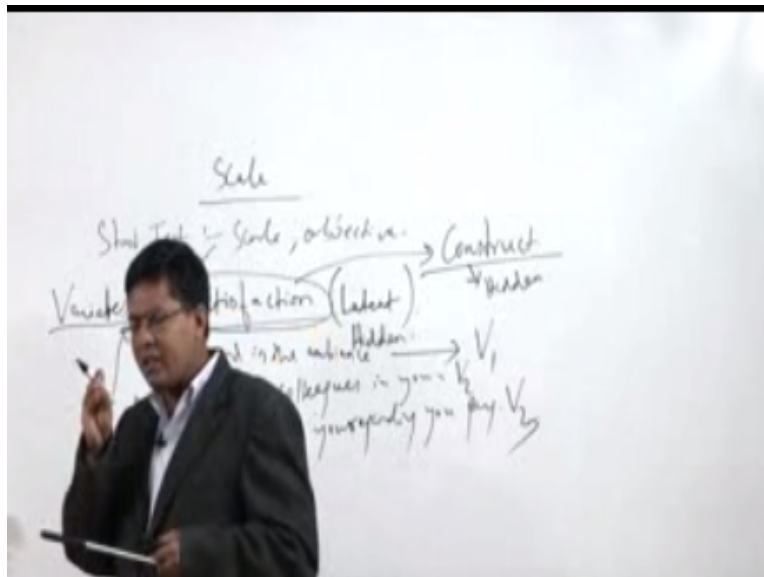
Your scale used in the study and the objective of the research right, so let us understand and the problem is that in basic research it is very simple to understand that you have units of measurement. But in terms of social science it is slightly complicated because many things are

not very clear. That is why it becomes little nebulous and it become hazy. So that is why we need to understand what the scale is? That means how to develop the scale in the social science okay, so let see this.

(Refer Slide Time: 01:42)
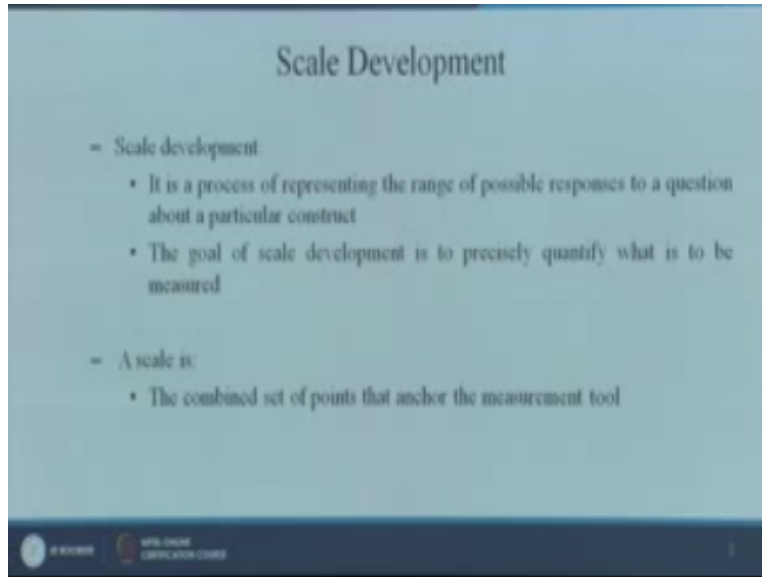
So what is scale development? Scale development as I said to conduct any research in social science, for example you have to conduct any study on job stratification for example let say. To conduct the study on job stratification you need to understand, you need to develop the scale for the term for example satisfaction may be all other connecting terms for it. So let see what is this exactly, it is a process scale development of representing a range of a possible responses to a question about a particular construct.

It says the goal of the scale development is to precisely quantify what is to be measured. So that means scale development helps to actually measure to calculate or to quantify, something that is very integral part of the research process for example, as I said satisfaction okay. Now the question is how do I measure satisfaction? To measure satisfaction I need to develop a skill right, so the scale because in many times in social science they are called latent variables right.

It could be the latent variable or the latent construct or anything. Now latent means from chemistry also understand, latent is something hidden basically. Which is not very clear on the surface, so it I something like hidden and obviously in social science because it is connected with the human emotion and all, belief and attitudes, many things are very much hidden and it is not

clear on the surface okay. What is this saying the scale is the combined set of points that anchor that measurement tool?

Now this is something similar to what do you do in a basic pure research, to calculate weight you need kg, so kg is one standard unit, similarly we need the step of points that can anchor the measurement tool or that can validate the tool and say this is something that can be used repeatedly again and again, so this is the validated basically. So let see these terms.

(Refer Slide Time: 04:01)
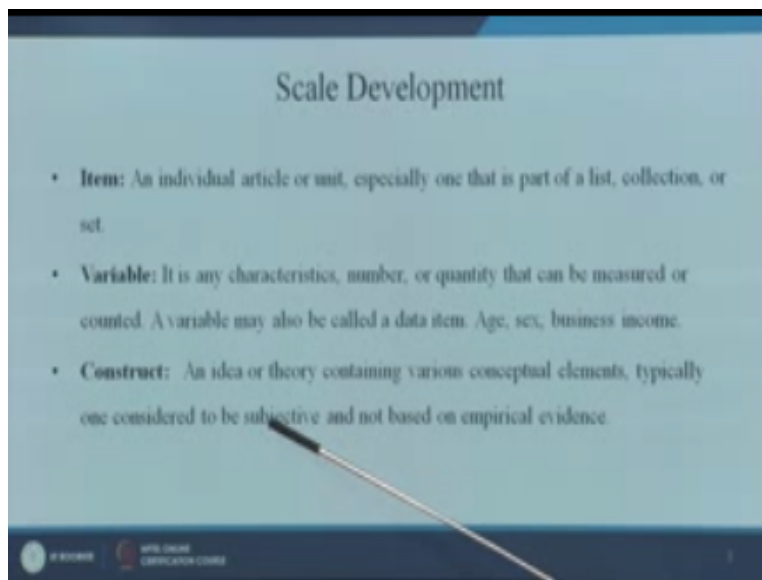
3 terms are very important, item, variable, construct right. An item is nothing but an individual article unit especially one that is part of list, collection of set. Now what do I mean by saying this, now let say. You are asking people about job satisfaction right, so when you are asking job satisfaction, you will have to ask not once but multiple number or things. That is related to satisfaction, for example how far or how good is the ambience okay.

Let say what is, how the people are, how are the colleagues in your office okay, similarly we can ask how happy you are right regarding your pay. So these are the questions these are basically the items, the pattern the units especially there is the part of list or collection of set. So entire collection regard doing on satisfaction, so each is an item that we have asked, when we do a question we will see that there are numbers of items in a questioner, which are used or each question can be described as an item.

Many times people confused this term item and variable right, although the difference is very thin in terms of social science research we will understand but still there is a difference, now what is the difference. Now variable is something as the name suggested it is something that variated. If the value changes, that means the value changes in which situation you can understand that right.

So it is the characteristics number or the quantity that can be measured or counted, a variable may be like a data item age, sex, business income anything right. so you can have categorical variables for example in case of nominal data you have categorical like male is 1, female is 0 or 2 something like that or you have the variable which is income, continues in nature 1000, 10000, 20000, 50000 right.

So when you are asking an item when there is an item and this item sometimes also normally what we do is we say this is variable 1, this is variable 2, this is variable 3 and it goes on. But understand the basic premise the basic definition of variable is something that varieties that changes basically. That is how the definition of variable is come. What is constructing? It is an idea or theory continuing various conceptual elements. Now it is the conceptual elements it could be abstract in the nature right.

Typically considered to be subjective and not based on the empirical evidence, now for example as I said, now satisfaction is nothing but a construct. A construct is something that is basically a amalgamation or combination of several items right. So now these are the items that are there to measure the construct satisfaction because if I ask you what is satisfaction it is a abstract term. If I ask you what is honesty it is the abstract term, . If I ask you what is loyalty it is the abstract term.

So a construct is basically something which is to understand it is number of items together to make it a construct right or you can sometimes say number of variables together to make it a construct okay. The right way of understanding is items. Now each construct as I say the construct is hidden, it is the latent structure. So in social science most of the constructs that we talk about are latent or hidden in nature.

That mean you cannot directly have the way of measuring it, so how do you measure? That question comes. So you need to measure with some other items with the help of some other items right, so this is how it goes right. How do you develop a scale?

(Refer Slide Time: 09:08)
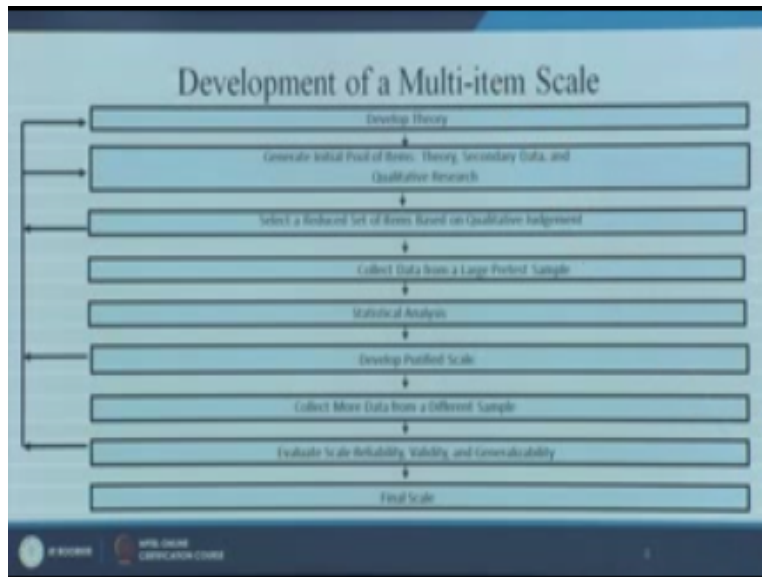
So the scale development this entire process is called a scale development, so when you develop a scale. So what you do is you start with a theory right, so what is the suppose you are conducting a study let say productive of employees or the in marketing you talk about consumer preferences or lifetime value of a customer or anything right. So first of all you have to start with the theory and generate the pool of the items please underline this, this is important.

You are generating the pool of the items, each item is a question right, which is from the theory or any other secondary data right and after this once you have collected these items. Now what is this item, these items have to have a strong co relation with the factor or the construct that you are measuring. That means these items actually represent the construct okay after this you do is once you have a large number of items, suppose to measure satisfaction we have got a large number of items.

Let say 10 items, now question is all 10 they might not be that applicable, so there what we do is a qualitative judgment or through some statistical measures, for example the exploratory factor that I will be explaining is the data reduction technique. We reduce the number of items and keep it to the minimum which explains the construct in better way, once we collect the data from a set

of pool of responded. Then next we try a statistical analysis, so that means is basically, initially. So we started with the theory right.
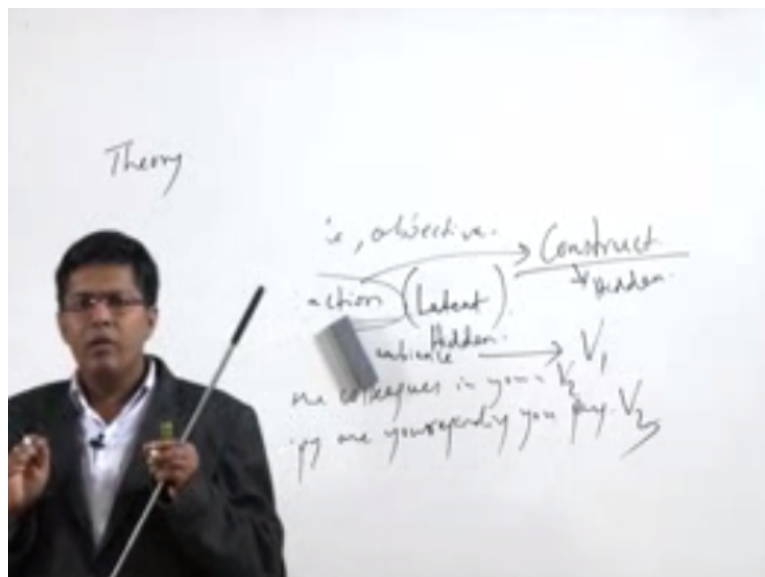
See please understand any research will only be correct obviously inductive research is different we know because we develop a theory. Here the deductive is where the most of the researches we are conducting in the application based are deductive. Here what we are doing is we have the theory on the basis of the theory, and then we are trying to test and say whether it is matching already existing theory or not okay. So theory we have and then we had the items and then we reduced the items may be.

And brought it to the limited number and then after that we tested it right, and the after testing right, first you collect data and after collecting the data here you test, there is something in between like which I have also said earlier you purify the data here right and then do it. So once

it is done you finally check for the reliability and validity two terms are important okay, when you come to making a scale or developing a scale which is very important, reliability and validity reverse.

We will learn it in the process right now okay in this session, so finally the scale is made okay. So let see.

(Refer Slide Time: 13:13)
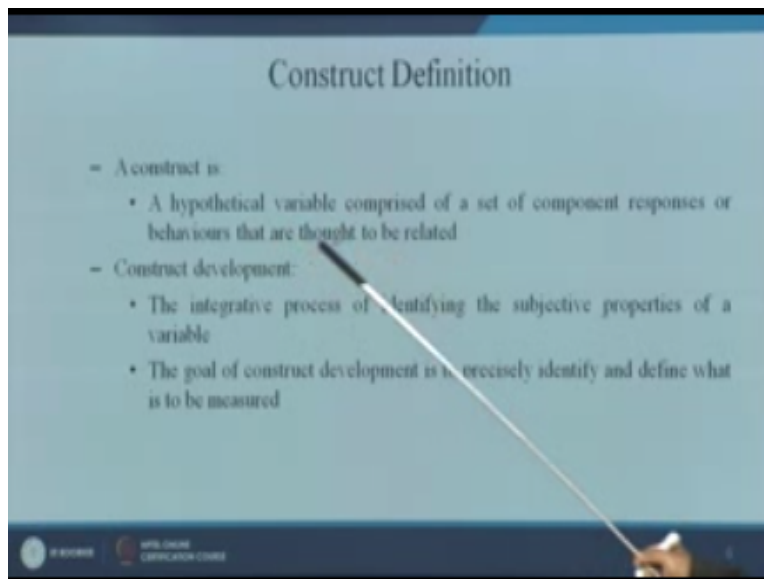
Different steps in the scale element we just spoke about it, construct has again defined here, a hypothetical variable comprise of a set of component response or behavior that are thought to be related please again this something I was saying, so that the items which I said they had some relationship with the let say the construct. It is the integrate process of identifying the subjective properties of a variable. So it is not that only we have a quantitative approach, we first start with the theory.

So we have the subjective understanding about it, the goal of the construct development is to precisely identify and what is to be measured. So it helps you to understand what is the item? That is to be measured so all are not required, when you are developing the items all of them are not might exactly defining the construct, which are the once we are required and they truly define and explain the meaning of the construct that is what we do.

(Refer Slide Time: 14:25)
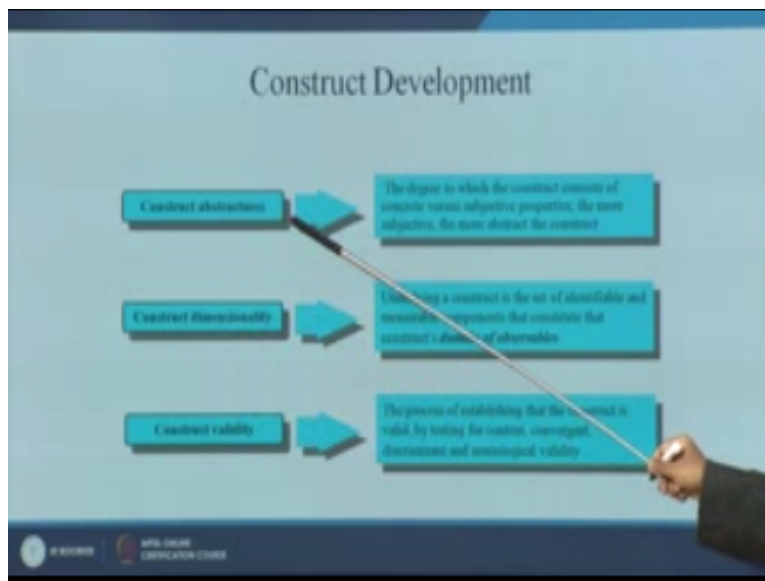
So first construct abstractness in the process of three things, construct abstractness, dimensionality and the validity. What is abstractness is, it say degree to which the construct consists of subject versus subjective properties, the more subjective the most abstract and social science exactly this is what happens. Dimensionality it say it set of identify and measurable components. Now in the search we need to have a scale, a scale should be a question or an item should be measurable right.

These are the characteristics of the items right that you are using. So that should also be there right. 3rd is it should be valid the construct should be valid, so there are several test which I will be explaining further. So it has to be validated, now what do you mean by validated, let ask the question valid. Whether a chalk is a valid instrument to write on this board not at all, so I should be writing on the help of marker right, now there can be another material also.

Some markers which cannot be rubbed or cleaned, are there permanent markers, should that be used in that board, is that the valid instrument no again. So the right valid instrument is only this, this kind of marker right. So you have to understand the scale that you are developing it should be valid and then it should be reliable right.

(Refer Slide Time: 16:14)
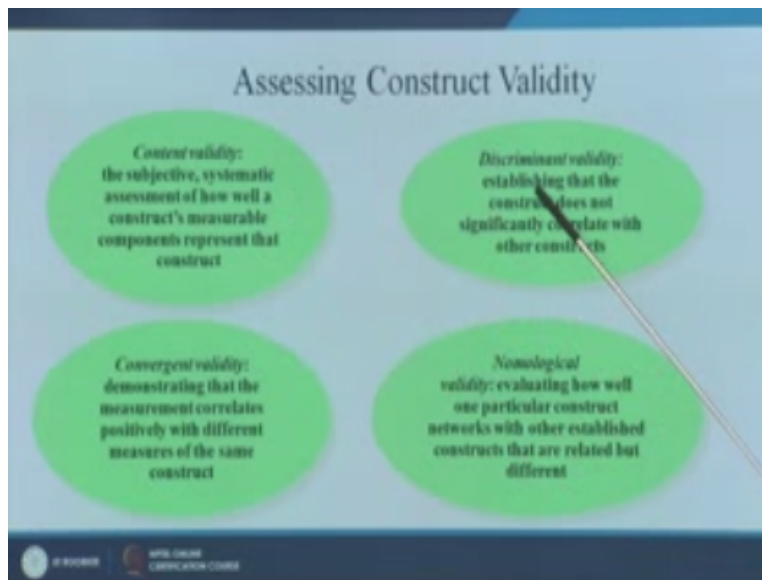
Constructive validity has got 4 parts right, content validity, and discriminative validity, convergent and nomological validity right. So basically this is the construct, so the construct was the satisfaction, let say one of the very popular construct that we use in marketing literature is trust. Trust is the most frequently used construct because it is used in every theory. So now what is content validity, it says the subjective, systematic assessment.

Of how well measurable reflects the components, the items that you are using are really will help you in measuring that construct yes or no, that is what you know content validity says, so it is from the content, now it is like the drama, movie or picture or any show. The content we say how good is that content is poor or bad still the concepts remains the same but then it does not define it properly in a nice manner right.

Similarly next let us go to this one nomological validity, now what is nomological validity? Nomological validity means whatever you are trying to prove that should be corresponding to the theory in large. Evaluating how well a particular construct networks with other which are related but different, so theoretically whatever as been proved, now whatever you have been proved, if it is coming true that yes x, y, z has said this and now we also come to a similar conclusion.
Then where we say that there nomological validity that is something that rhymes with the theory right if it is not then we say it is against. But if it not then we say it is against, but there is something which I have my own interpretation here. Nomological validity is important but this is my personal opinion all the time if you follow the nomological validity, there is a very less chance in defining the new area of research are trying to find something new.

For a entirely new thing nomological validity might not fit properly okay. So next is convergent validity now what is convergent validity, so if you can understand from the term, something like this is it is converging. So are converging to the construct, now what is to be converging that is the item okay. Demonstrating the measurement correlates positively with same construct is as I said are item1, 2 and 3.
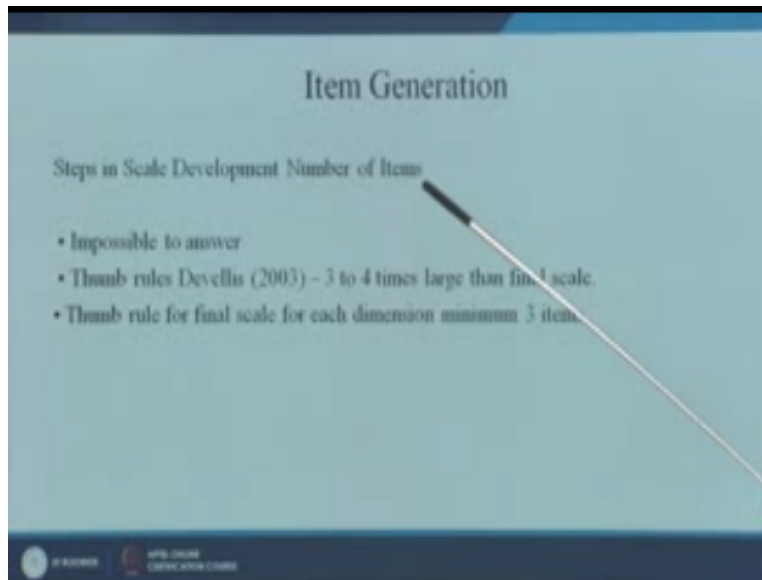
Now what is the convergent validity are they converging means is there a correlation between them, we will say they all are moving towards to the same direction that means they purpose they are trying to do exactly what they were made for. But suppose there is the poor correlation let say, two of this items right then what happens, we will understand that one of them is at least, if there is no co relation that means they are not exactly defining the construct.

One of them have got less relevance it is the mistake and we need to find out why this correlation is poor, the members of the family, if it is from a one family, then there as to be high correlation between the members, if there is a poor co relation members of the one family, that means we have make our understanding clear that they might not be from that family, he might be from other family and this might be the mistake from our side to understand that.

Discriminative validity is something where it says that one construct should be necessarily different from other construct, what it says establishing that the construct does not significantly co relates with two other constructs okay. Suppose your construct is honesty and trust worthy, now sometimes honesty or loyalty let say, now sometimes it looks at to similar to us, if they are very similar then point is why should we have two different constructs we can have only one constructs.

Then the items within the constructs they would also have the strong correlation among them but that happens we are fundamentally wrong because that means there is no need of two constructs. So here what happens we will check for the correlation between the constructs right. Correlation between the constructs between the constructs separately very poor correlation that should not be co related it should be as far away for each other. So items generation for examples steps in scale element the number of items.
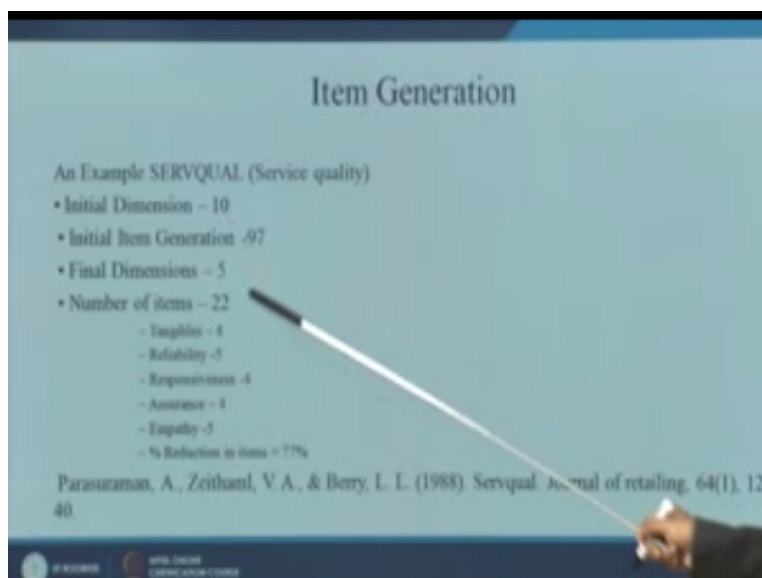
(Refer Slide Time: 21:53)

Item Generation

Steps in Scale Development Number of Items

- Impossible to answer
- Thumb rules Devellis (2003) – 3 to 4 times large than final scale.
- Thumb rule for final scale for each dimension minimum 3 item.

How many items should be generated? How many question should be generated, this is impossible to answer but the thumb rule says 3 to 4 times larger than the final scale. So let me tell you, suppose you have let say you can simply understand for each construct have at least minimum of 3 items, if you have 3 items than obviously you know the explanation power will get reduced okay.

Although sometimes 2 items have been used you do not get confused because there is a 2 items which the construct as been made but that is not a favorable way okay.

(Refer Slide Time: 22:54)

Item Generation

An Example SERVQUAL (Service quality)
- Initial Dimension – 10
- Initial Item Generation -97
- Final Dimensions – 5
- Number of items – 22
    - Tangibles - 4
    - Reliability -5
    - Responsiveness -4
    - Assurance - 4
    - Empathy -5
    - % Reduction in items – 77%

Parasuraman, A., Zeithaml, V A., & Berry, L. L. (1988) Servqual. Journal of retailing, 64(1), 12-40.

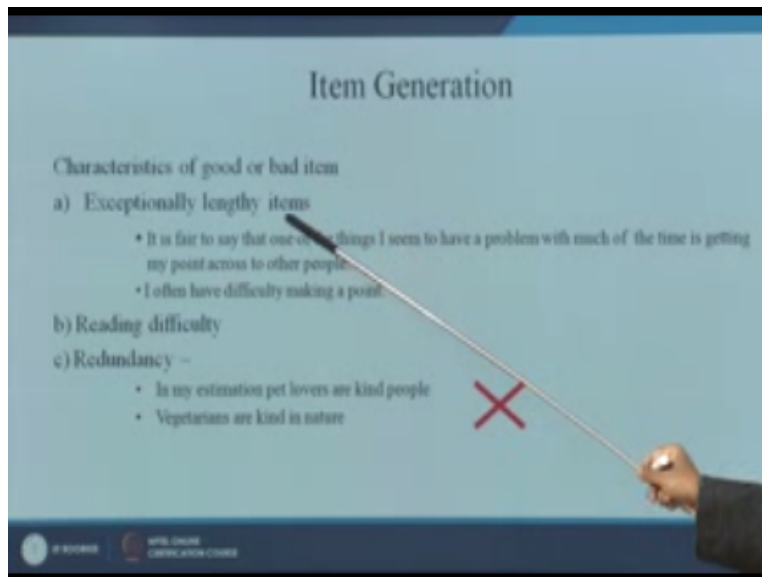Now this is one a very popular SERQUAL service quality example which I have taken, so initially if you see.
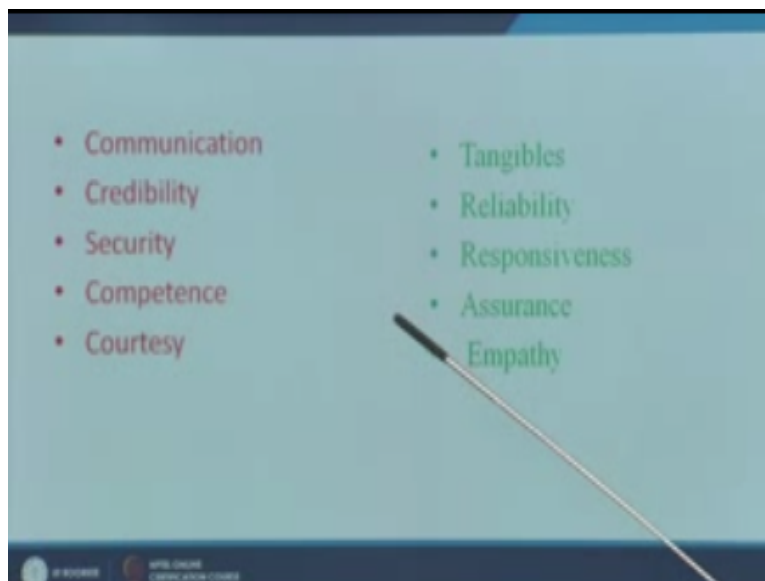
Initially there were 10 dimensions 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, now in the process we did once you identify there were 97 items for this 10 right. Finally only 5 dimensions were maintained why that only5 is dimensions were maintained? Because the next 5 were not explaining well to the requirement of the researcher, number of items that were maintained as only 22 and what were the 5 items, now these 5. And these 5 are the one which was removed right.

And each item had these number of had again 4, 5, 4, 4, 5 items. Now I hope you understand that if you have suppose three measurement constructs in the study are 4. Each constructs at least 3 to 4 items, so that means you need to have 12 question as good as that right. Forgetting the demographic part please, do not add the democratic part into it specifically it to the construct.

(Refer Slide Time: 24:23)

Characteristics have good item, item should not be exceptionally lengthy reading should not be difficult, although we will do this in the next follow of the session like in the questioner, some of the items are redundant so it should not happen right.

(Refer Slide Time: 24:43)

Negative question should not be there, it should not be double barreled we will see some.

(Refer Slide Time: 24:50)

Validity

- The **validity** of a scale may be defined as the extent to which differences in observed scale scores reflect true differences among objects on the characteristic being measured, rather than systematic or random error. Perfect validity requires that there be no measurement error ($X_O = X_T$, $X_R = 0$, $X_S = 0$).
- **Content validity** is a subjective but systematic evaluation of how well the content of a scale represents the measurement task at hand.
- **Criterion validity** reflects whether a scale performs as expected in relation to other variables selected (criterion variables) as meaningful criteria.

Now coming to this important part that is validity, so as I said the validity the instrument as to be valid this is what the example what I gave right. Now in a perfect validity there should not be no measurement errors that means measurement errors is like something there should not be no systematic or random error. Random errors are basically something like fluctuation in persons current mood it is random. Misreading or misunderstanding he has not understood this well, measurement off the individuals on different days or in different places, these are the errors happen in random error.

But systematic error will happen when the style of measurement towards there is the example sources of error including the style of measurement, tendency towards self promotion, cooperative reporting. And other conceptual variables so systematic error and random error should be minimum okay. Now let us go the content validity, content validity is the subjective but the systematic evaluation of how well the content of represents the measurement task we have discussed it. Criterion validity is as a scale performs as expected in relation to other variables as meaning full criteria this is important.

Now criteria you are saying that this is used for some timer say criteria varies of two types again, that means for example suppose I am good student I have the good score right, let say in exam, it is somebody can predict that I will do well in the MS or the course that I have joined it right. So this is why criteria validity helps you forecast the future in one way, you can understand that right. it reflects whether the scale performs as expected.
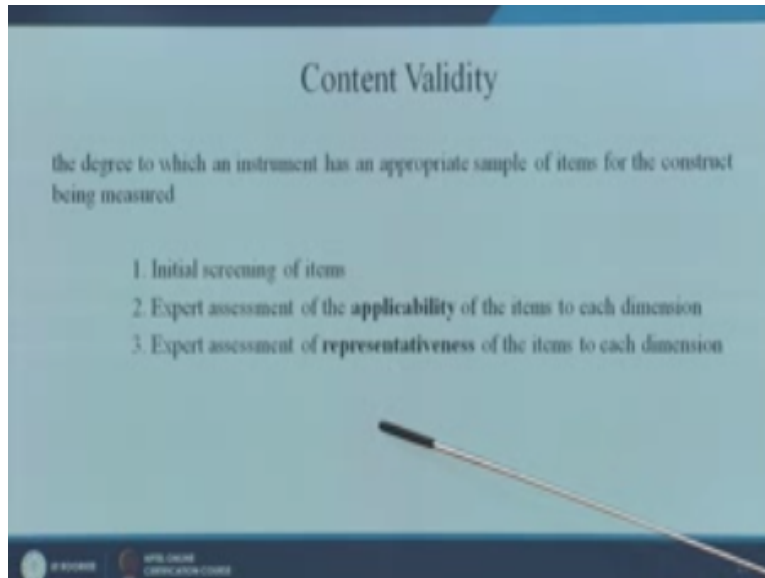
(Refer Slide Time: 27:05)
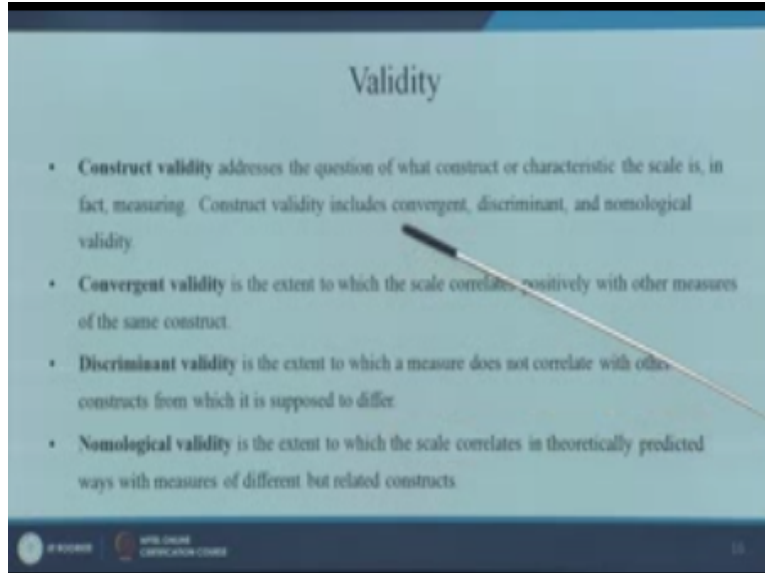
Construct validity what I explain, Construct validity is actually something which is combination of these three rights so the construct validity includes convergent validity so where the item should be converging so there should be high degree of co relation among the items. Nomological as to explain theoretical and what you get that should be similar, discriminative validity I have explained through the constructs right.
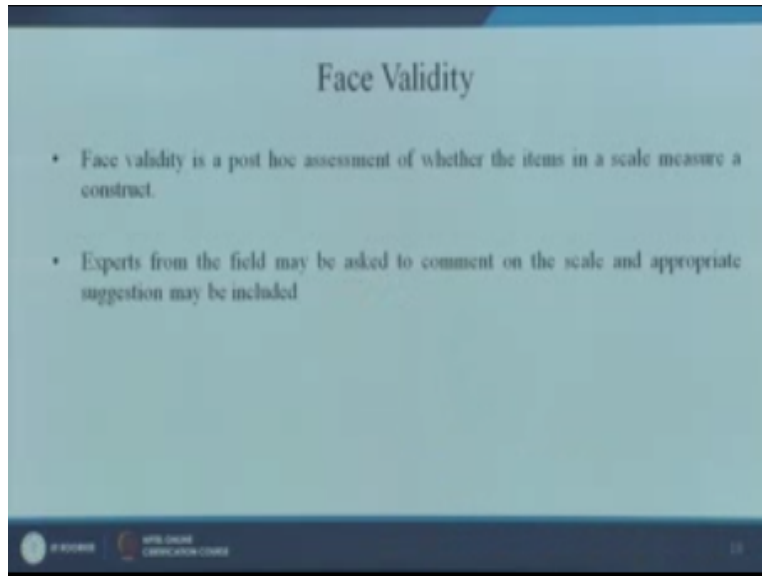
Now if you can see that extent to which a measure does not correlate with the other constructs, so it is not between the items but it is between the constructs okay.

(Refer Slide Time: 27:41)

Content Validity

the degree to which an instrument has an appropriate sample of items for the construct being measured

1. Initial screening of items
2. Expert assessment of the **applicability** of the items to each dimension
3. Expert assessment of **representativeness** of the items to each dimension

So content validity is done through the expert assessment where the applicability, whether the item is actually representing the construct or not it is to be done okay.

(Refer Slide Time: 28:00)

Face Validity

- Face validity is a post hoc assessment of whether the items in a scale measure a construct.

- Experts from the field may be asked to comment on the scale and appropriate suggestion may be included

Face value also similar to that we measure whether the items in the scale actually they are measuring the context, so here also we do expert only opinion so sometimes there is the line of difference very thin so content and expert are many times taken to be similar right.

Reliability is something now very interestingly I will show you the diagram. What is the reliability first?

(Refer Slide Time: 28:26)
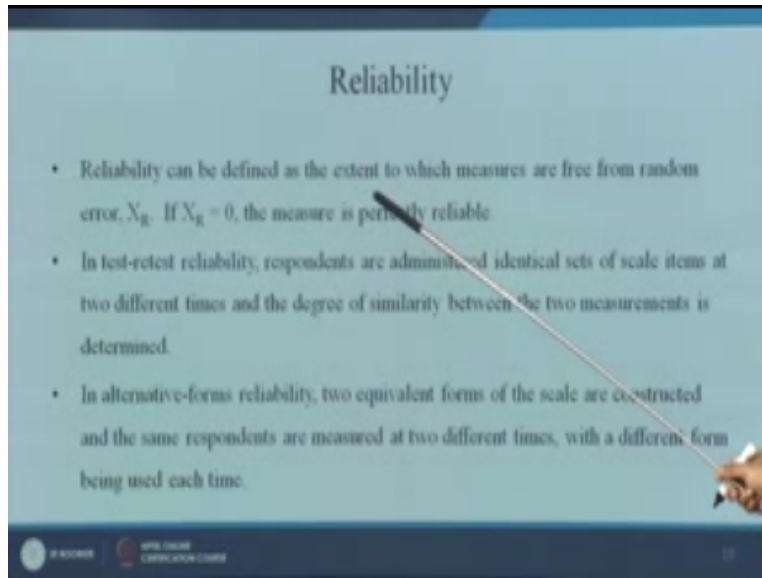
Reliability is something where you are saying that this pen is for example, first of all we will find out it is valid. Now it is reliable, now let us take another example otherwise, you must have checked your weight correct, at different places when you check your weight on a machine and it says the more or less same weight then you say the machine is reliable but suppose there I the variation in the score right then you say that this machine is not reliable.

The point is giving the same value or trying to measure the thing again and what is intended if it is giving the same score repeatedly then we say that is reliable instrument okay. Now you see it says, it is extent to which the measures are free from random error and the measure is perfectly reliable right, there should not be any random error, for example on the machine where you are standing on measuring the weight. There should not be any random error it should be giving the same kind of random error.
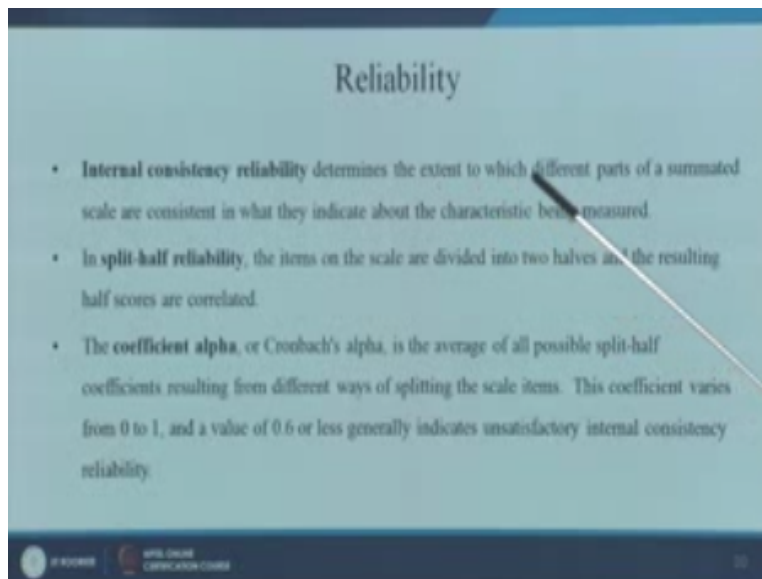
If there is systematic error we can still handle it right, so what is happening here the different types of reliability testing for example test it, again test it. So when you test it two different times look at their values, if there is not much difference then you say it is reliable thing. Alternative function reliability, two equivalent forms of scale are constructed and the same responded are measured at two different times.

Now these are two different forms please do not confused with the meaning, here two identical state of scale items at two different times and the degree of similarity between two measurements in determined and here what we are doing is responded with two different times. So something

like you have one test and then you will make another test maybe another method and you will look at the scores or you have the same method at two different times, 3 times and check the score.

So if you find any difference there then there is the error, so these errors are not advisable or should not be there okay, internal consistency is something like.

(Refer Slide Time: 30:42)

The different parts of a submitted scale are consistent in which they indicate, now forget this simple terms if you want to understand is internal consistency will only be high please remember when there is a high correlation among the items, then only there will be internal consistency because the submitted scale will again say that means this construct will.
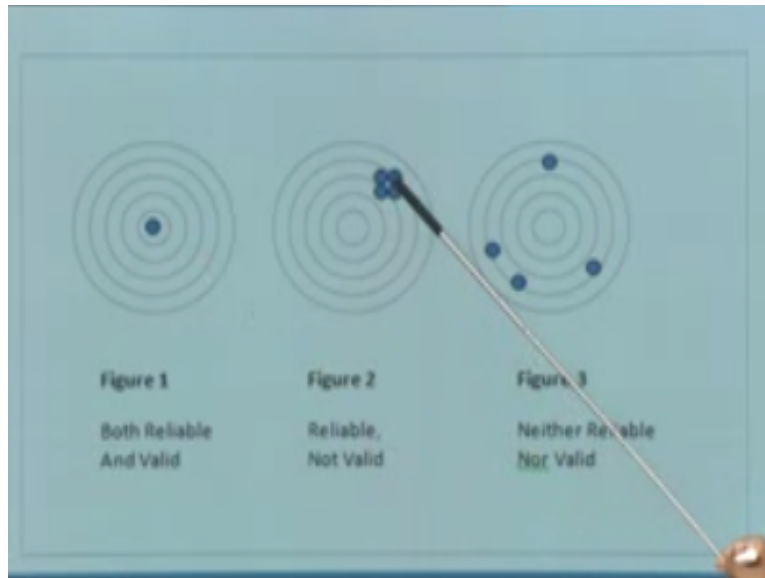(Refer Slide Time: 31:04)

Whatever result is construct will give if you do it on a individual like a simple regression or a multiple regression, so instead of doing a multiple regression if you do individual simple regression and whatever values you get and you do one multiple regression and whatever values you get is the same then we will say more or less they are giving the same result okay. split half is something where the data is divide into two parts and then each part is checked and then compared the results are compared with a similar then say there is a reliability, okay into two halves and the resulting half scores are correlated. Coefficient alpha Cronphach's alpha is a, is basically nothing but an average of all possible split half techniques or coefficients that has been generated.

So it is an average now it is said that for social science it has to be when I said it is 0.7 with a popular paper was Nonaly's and but you can say anywhere around 0.6 or above can be consider a good amount of reliability but please remember again and again I am repeating the basic you know of which all these happens is nothing but the correlation among the items okay.

(Refer Slide Time: 32:16)

Figure 1 — Both Reliable And Valid   Figure 2 — Reliable, Not Valid   Figure 3 — Neither Reliable Nor Valid

So it says both reliable and valid this is reliable the machine it is falling within the a it is reliable but you know sorry it reliable but not valid it should be heating somewhere here but it is not it is at one point although he data what it is giving is more or less same sop that is why they are reliable but they are not valid third case it is saying neither reliable nor valid neither it is heating to the target and not they are even giving the same score so it is neither reliable not valid okay. Finally we do a pre testing pre testing is usually done to check whether our instrument is correct or not correct.

And here while doing this is again you know confusion the mind of researchers what should be the sample size for a pre testing generally take the sample size as 10% right so if you have 300 so 30 could be your sample size for a pre testing or pilot testing okay. But pilot testing is very important because it like understands that what I am doing is correct my process is appropriate okay. So these are all that I have explained to you right and this is the diagrammatical form only.

(Refer Slide Time: 33:25)
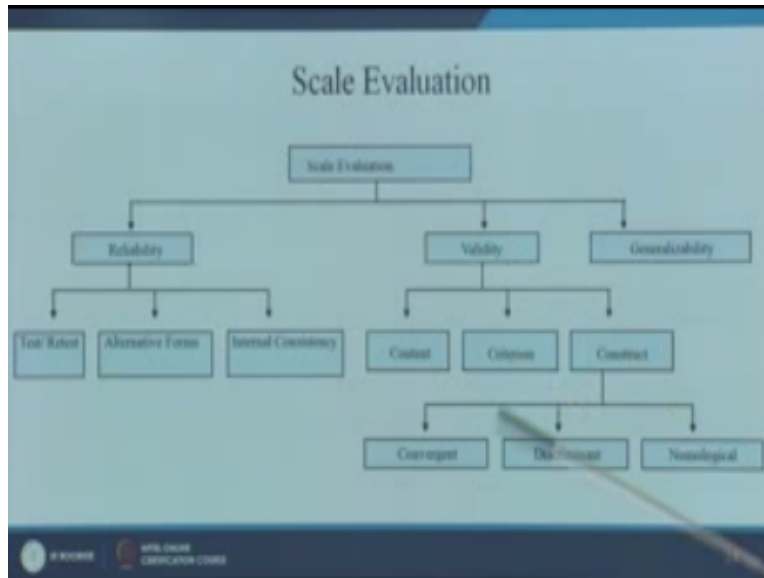
You can later on also check it right, what is this is some of the potential sources for error so they errors can happen because of like personal factor situational factors lack of clarity of the scale mechanical factors like poor painting right to many items club together right, or overcrowding sop these are some of the and how the administrator as effectively the interviewers as conducted in the study all these things are different points through which the scale can some errors.

One has to be very clear reducing this errors so finally when we you know have recap of the session is like what is the scale what is then item basically what is the variable what is the construct right what is the basis of the relationship which is correlation is said right and how does a construct what is actually construct it is hidden or latent in nature and how do you measure it, what should be the reality value something on 0.6 or above right which good enough in social science.

If you go to an non social science then it has to be much higher this is again one more difference which I did not say for an explanatory condition it is 0.6 but for it is conformity suppose then the conformity factor which will be taking sometime later it has to be little higher obviously because it is already a defined scale which is already proven. So these all things you need to understand and how you will have to do a pilot test reduce the number of items have a large number of items reduce them then test make a pilot test and then finally come to a conclusion and then use this scale validate and check this scale for reliability and then this is becomes a universally excepted

scale and this scale can be then next used anywhere right..So this is all for this section, thank you so much.