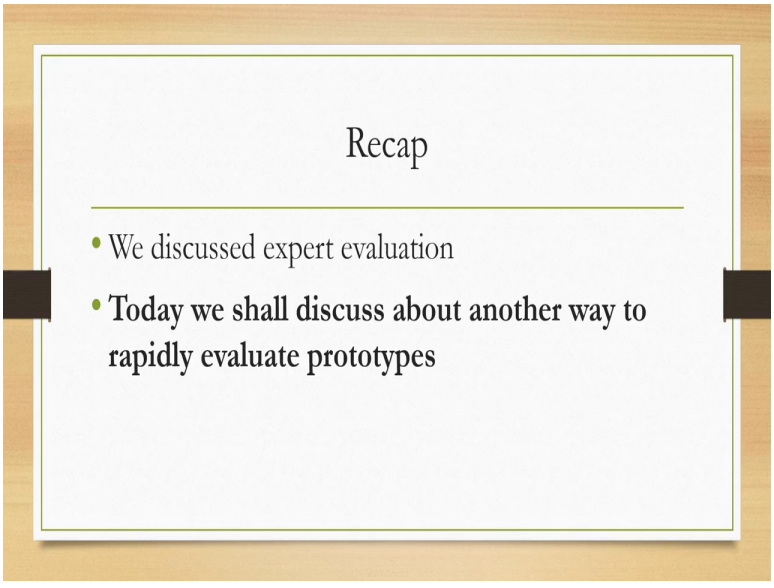
Design and Implementation of Human – Computer Interfaces
Prof. Dr. Samit Bhattacharya
Department of Computer Science & Engineering
Indian Institute of Technology, Guwahati

Lecture: 17
Prototype Evaluation II

Hello and welcome to the NPTEL MOOC's course on design and implementation of human computer interfaces lecture number 16 where we will continue our discussion on evaluation of prototypes. So, in the previous lecture we started our discussion on how to evaluate prototypes. So, the constraint that we have is that once a prototype is created it has to be evaluated quickly. Here we are assuming that the design building of prototype and evaluation of prototype forms a cycle.

And this cycle repeats frequently in the early stages of design till we arrive at a design which is having very less significant amount of usability issues so it is very important to be able to evaluate the prototypes quickly and rapidly. In the earlier lecture we have discussed about specific methods of evaluations in particular we learned about expert evaluation methods two methods we discussed one is cognitive walk through other one is heuristic evaluation.

(Refer Slide Time: 02:12)



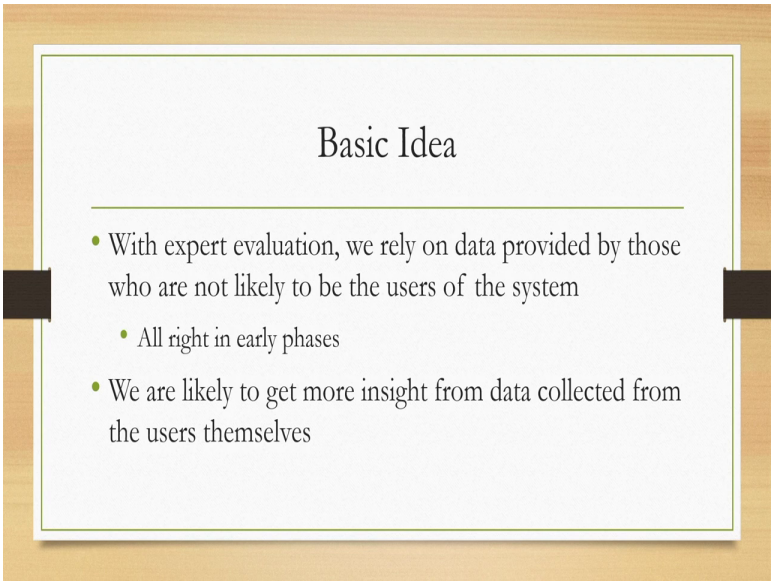
In this lecture we will continue our discussion on how to quickly evaluate prototypes. So, today we are going to start our discussion on another way of evaluating prototypes rapidly which is

user evaluation. So, here we will be discussing methods to involve end users in rapid evaluation of prototype. Recollect that in our earlier discussion on prototype evaluation we said that it is expert evaluation that means end users involvement is not necessary.

Although optionally they may be added in the evaluation team however today what we are going to discuss in this lecture is based on user evaluation. Now generally evaluating prototypes or any systems with users is time consuming process involving elaborate setup detailed data collection and analysis methods which entails cost and time. However there are ways to quickly evaluate systems or prototypes with users although the outcome will not be as robust as what we will get if we follow a rigorous evaluation method.

But still in the early phase of design such a quick evaluation would be beneficial. So, let us see how we can do that.

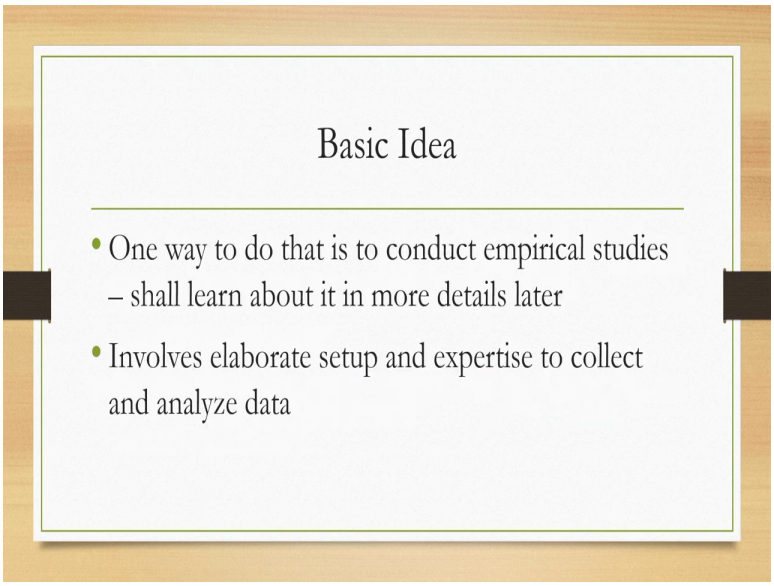
(Refer Slide Time: 03:41)



So, with expert evaluation that we have discussed earlier we rely on data provided by those who are not likely to be the users of the system. Now, that is one thing that we emphasize that expert evaluation primarily relies on expertise or domain expertise of evaluators. Now in the early phases we make use of that which can be considered to be a replacement of evaluation by end users.

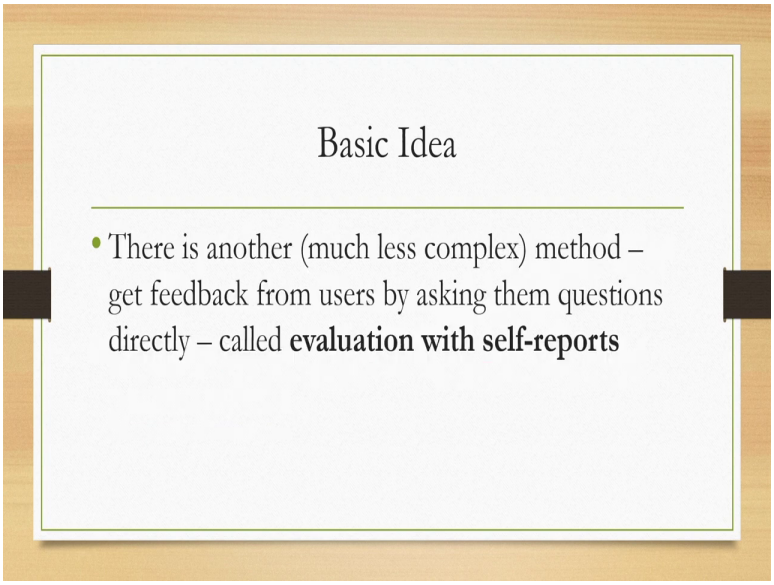
However if it is possible to involve end users then we are likely to get more insight about usability issues of a particular product from the data collected from such users.But we have to do it in a rapid fashion.

(Refer Slide Time: 04:33)



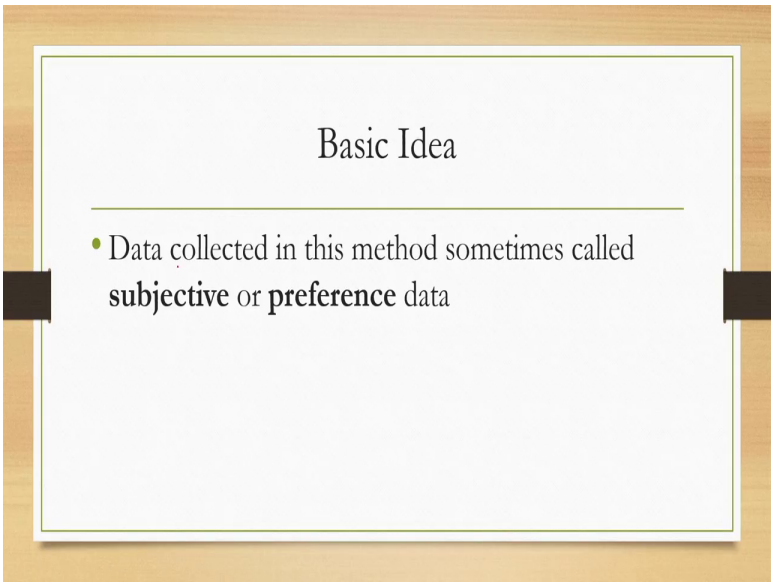
Now one way to do that as we mentioned as one stage of the interactive system development life cycle one way to do that is through empirical studies. Now that is actually a very rigorous process and we will learn it in more details in a later lecture. This involves as I said elaborate setup and expertise, expertise of the experimenter as well as expertise for analysis of data to collect and analyze data which is not an easy thing to do and it takes time.

(Refer Slide Time: 05:15)



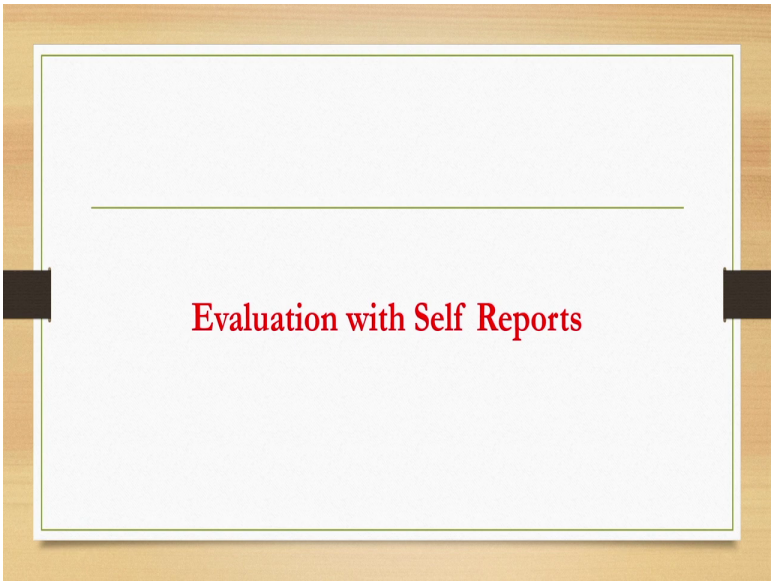
Now there is another way to collect end user feedback on a design which is much less complex and here we can get feedback from users by asking them questions directly. So, you ask some questions and answer to those questions are likely to give you some feedback about your design idea this is called evaluation with self reports. So, this is the method that we are going to discuss in this lecture.

(Refer Slide Time: 05:56)



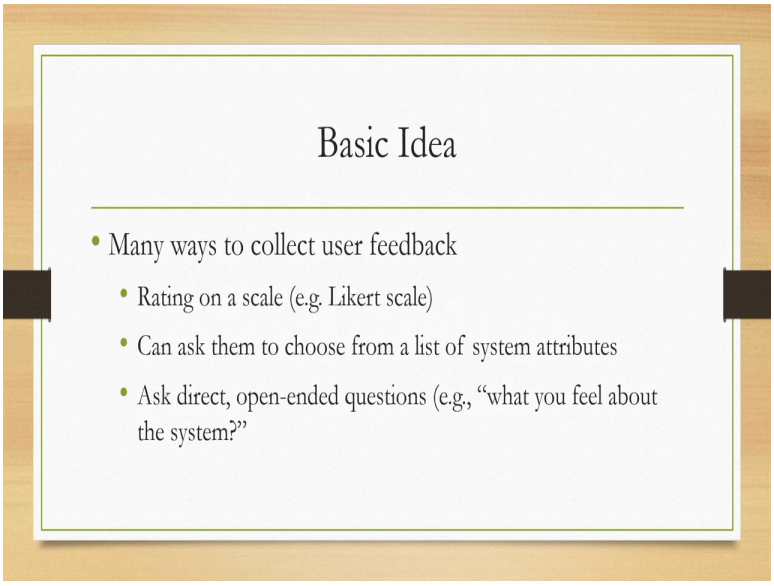
The data that we can collect through this method can sometimes be called subjective or preference data that is another way of referring to the data that we collect through this method.

(Refer Slide Time: 06:14)



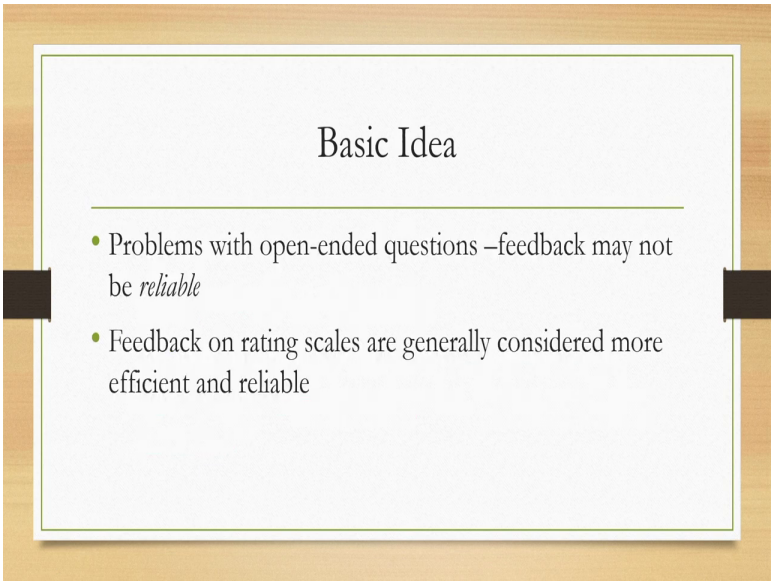
Now let us learn in details what is evaluation with self reports.

(Refer Slide Time: 06:20)



Now when we want to get some data about usability issues of a product how we can get the feedback from end users. So, there are different ways possible. We can ask the users to provide some rating on a scale a rating scale a typical example is a Likart scale for providing feedback. We can ask them to choose from a list of system attributes and ask for their feedback about those attributes. Also, we can ask them direct open-ended questions such as what you feel about the system. So, there are different ways to get the feedback.

(Refer Slide Time: 07:09)

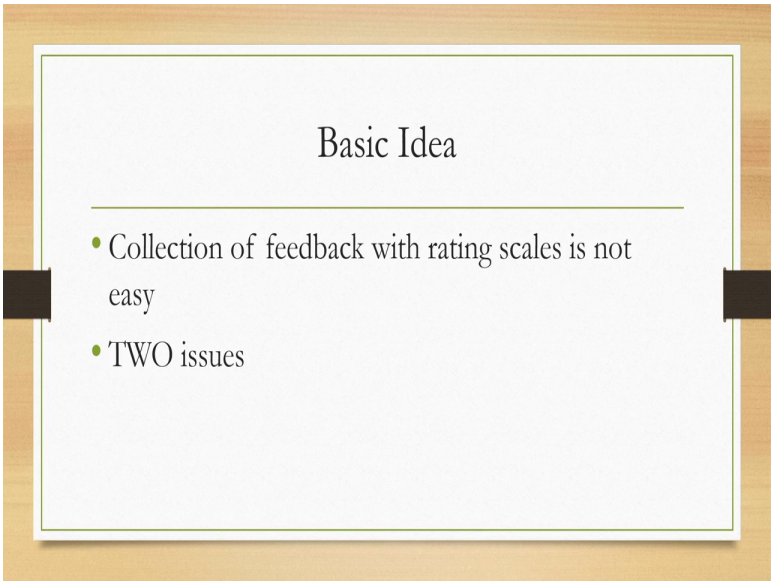


Now which way we should use? Intuitively it may seem very easy to do if we simply ask them about their feedback through open-ended questions like what you think about the system how do you feel about the system do you think there are some issues and similar such open-ended questions but as you can understand if I simply ask some user to give feedback based on open-ended questions.

The results will not be very reliable because at that point of time when the user is providing the feedback several things influences the answer including prior exposure to the system mood of the user at that point of time so on and so forth. Now there is no way to actually analyze objectively the reliability of the answers based on open-ended questions. In contrast if we try to take feedback based on rating scales such as a Likart scale then these type of feedbacks are generally considered to be more efficient and reliable.

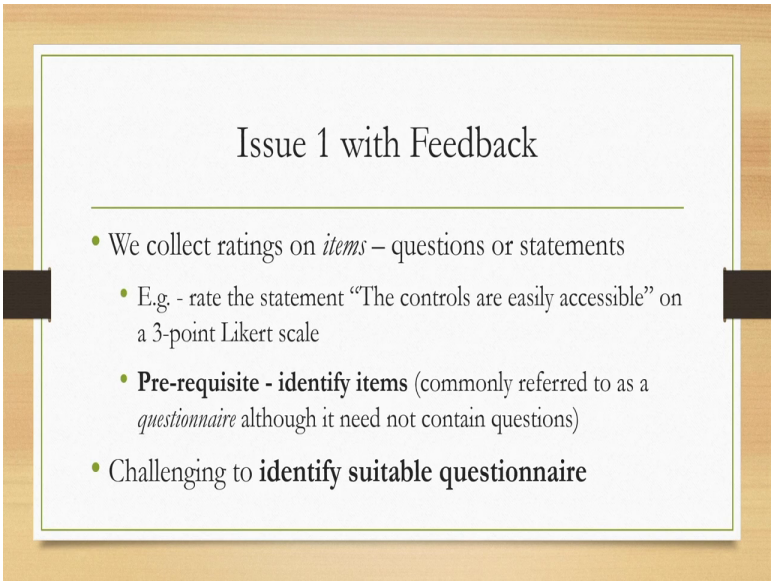
So, instead of going to collect feedback based on open-ended questions ideally we should aim for collecting such feedbacks using some sort of rating skills such as a Likart scale.

(Refer Slide Time: 08:50)



But that is also not easy. So, when we are planning to collect feedback using a rating scale it is not an easy thing to do there are broadly two issues involved here what are those issues.

(Refer Slide Time: 09:05)

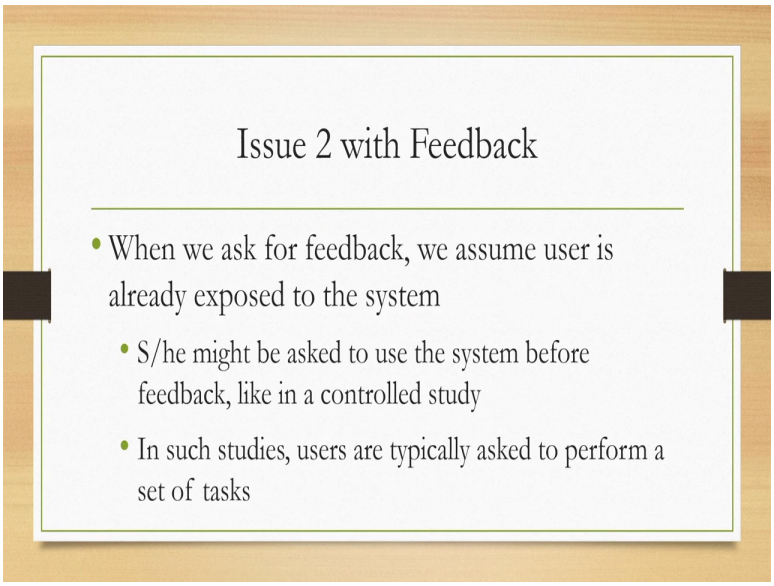


So, when you say we want to collect rating about some interface or rather design of some interface how do we do that we collect it based on some questions. These are generally called items which are typically can be in the form of a question or in the form of statements. For example there can be a statement about some design that the controls are easily accessible. Now on this statement we may ask the end users to provide some rating on a three point Likart scale.

Where one may refer to highly agree two may refer to neutral and three may effort to disagree. So, this type of statements can be used to collect the rating. So, what it indicates it indicates that there is a prerequisite before we go for using some rating scale based feedback mechanism. Here what we need we first need to identify the items. Now these items can be in the form of questions or statements as I just mentioned.

However when we have a set of such items this is typically referred to as a questionnaire although it need not contain questions it can be a set of statements but the general term used is questionnaire or items. So, it is very challenging to identify suitable questionnaire to collect the feedback that is one of the major issues in rating based feedback mechanism.

(Refer Slide Time: 10:59)

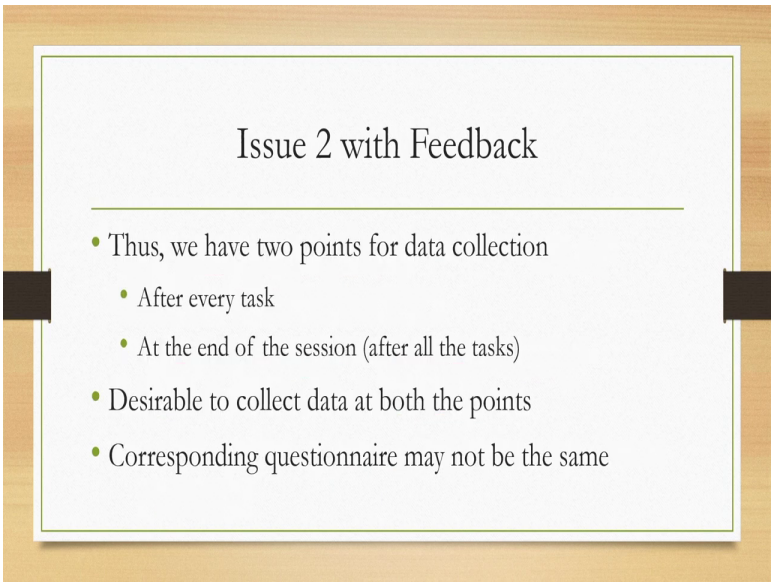


There is another issue when we ask for feedback we assume user is already exposed to the system otherwise the user will not be able to provide the ratings on the items in the questionnaire. Now how we can make it possible how the user can be exposed to the system she or he may be asked to use the system before feedback like in a controlled study. So, we can set up a controlled study and ask the users to use the system in that controlled environment.

And typically what happens in such studies is that users are typically asked to perform a set of tasks. So, a set of tasks is provided to the users and they are asked to perform those tasks using

the system in the controlled environment. So, that is generally how we can expose the user to the system.

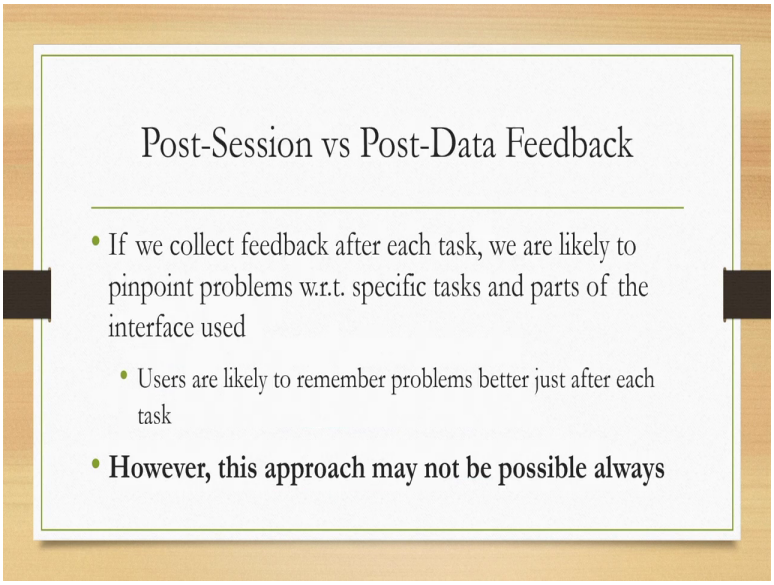
(Refer Slide Time: 12:13)



Now there are two points where we can collect feedback one is after every task remember that in a controlled experiment we ask the user to perform a set of tasks. So, after every task we may try to get the feedback of the user about the system or after all the tasks are performed which is generally called a session at the end of the session we can try to get the feedback. Now if we decide to collect feedback at one point and ignore the other point that is generally not very advisable thing unless there is some real problem in collecting data.

So, it is desirable to collect data at both the points after each task and at the end of the session but when we are using questionnaires to collect feedback. The questionnaire for getting feedback after each task and the questionnaire for getting feedback after the whole session need not be the same. So, there may be differences corresponding questionnaire may not be the same that we have to keep in mind so accordingly we have to design the questionnaire.

(Refer Slide Time: 13:30)

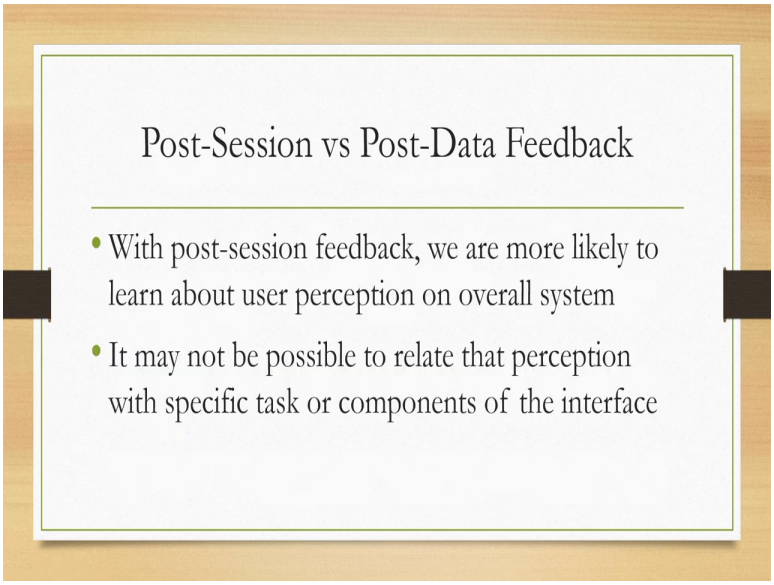


Now which one is better is it advisable to go for collecting feedback after each task or after each session. So, let us try to quickly compare these two data collection points one is post session versus post data feedback. So, if we collect feedback after each task we are likely to pinpoint problems with respect to specific tasks and parts of the interface used to carry out the task because users are likely to remember problems better just after each task.

So, if we have faced any issue while using the system for carrying out a specific task that issue is likely to be remembered by us and expressed by us if we are asked to provide feedback just after the task is completed. Instead if we are asked to provide feedback after end of all the tasks then it is possible that we may forget some of the issues that we have encountered during tasks that have been performed earlier.

So, this is the advantage of post task feedback over post session feedback but this approach may not be possible always.

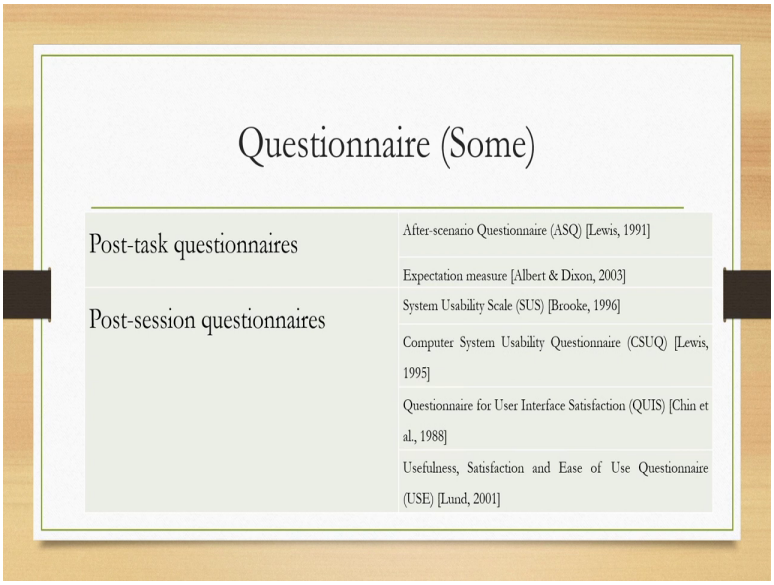
(Refer Slide Time: 14:58)



With post session feedback what we gain we are more likely to learn about user perception on overall system. So, after each task if we are taking feedback then we are likely to get issues related to execution of that task with the specific parts of the interface which were required to carry out the task. However it will not give us an overall idea of what are the issues. Whereas if we are collecting feedback after the entire session is over that means all the tasks are carried out then it is more likely that we will get user feedback on general issues related to usability.

So, we are likely to get general perception of the user about the overall system rather than only parts of the systems involved in specific tasks. Now it may not be possible to relate that perception with specific task or components of the interface. So, post session feedback need not be related to whatever we experienced after each task.

(Refer Slide Time: 16:25)

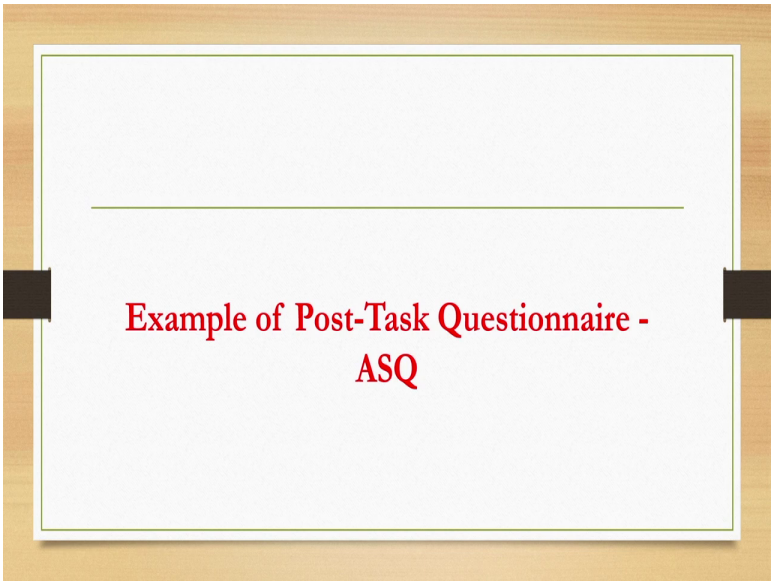


Now as I mentioned earlier so when we are trying to take feedback after each task vis-a-vis when we are trying to take feedback after the entire session or all the tasks performed we need to have different questionnaires same questionnaire will not work in both the situations in most cases. Accordingly different questionnaires have been developed over time to collect feedback at these different points of interaction with the system post task and post session.

So, let us quickly learn about few such questionnaires for different feedback for post task questionnaire several popular questionnaires are available like After Scenario Questionnaire or ASQ proposed by Lewis in 1991, Expectation Measure proposed by Albert and Dixon in 2003. So, these two are quite popular questionnaires meant for collecting feedback post task. Similarly there are several popular questionnaires designed to collect feedback post session which include system usability scale or SUQs questionnaire proposed by Brook way back in 1996.

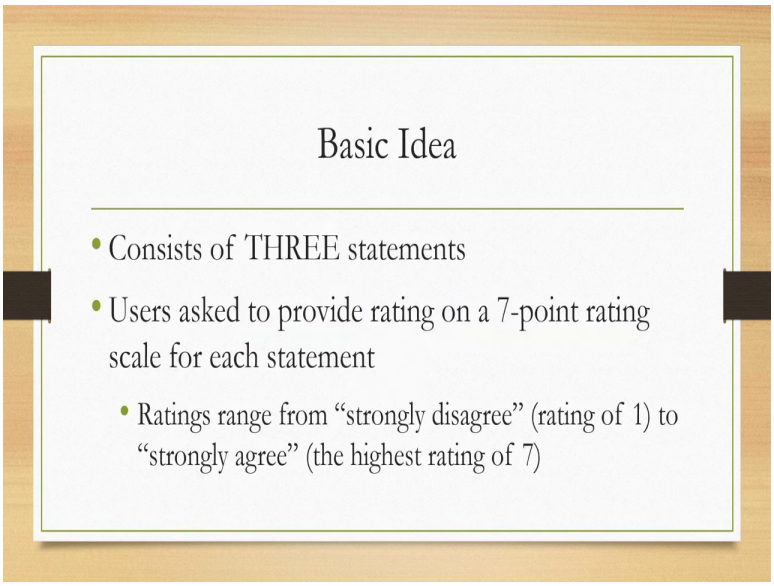
Computer system usability questionnaire or CSUQ proposed by Lewis in 1995, Questionnaire For User Interface Satisfaction or QUIS Chin et al proposed this in 1988, Usefulness Satisfaction and Ease of Use Questionnaire abbreviated as USE proposed by Lund in 2001. As you can see all of these are quite old primarily meant to collect feedback about GUIs mostly. And these are quite popular questionnaires still in use for collecting feedback at different points of interaction post task or post session.

(Refer Slide Time: 18:53)



Let us now try to understand each of these questionnaires with examples from each of these set of standard and popular questionnaires. So, we will start with post task questionnaire and the example that we are going to use is ASQ.

(Refer Slide Time: 19:13)

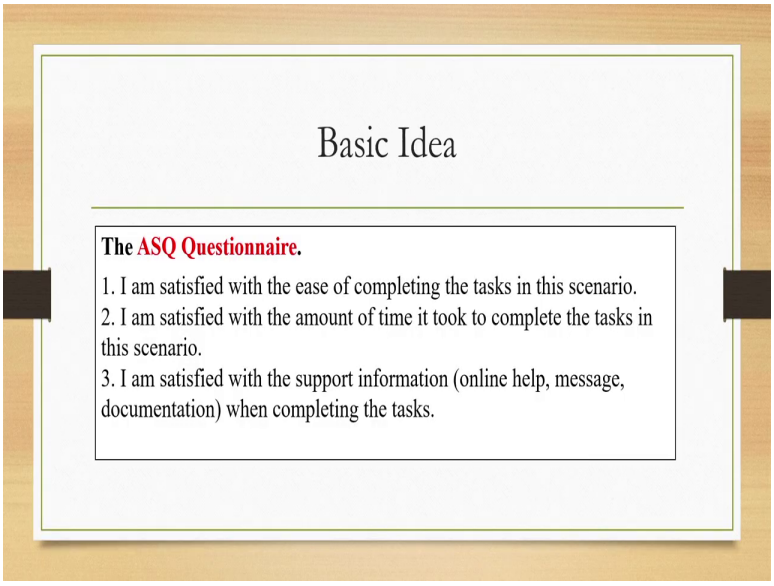


Now ASQ as I just mentioned ASQ is one questionnaire designed to collect user feedback after each task is carried out in a controlled experiment. Now ASQ contains three statements users are asked to provide rating on a seven point rating scale for each statement. So, there are three statements and on each statement users are expected to provide a rating on a seven point rating

scale. Now ratings range from strongly disagree which is the rating of one to strongly agree which is the highest rating of seven.

So, after every task user has to fill up the ratings for each question where the ratings vary from strongly agree to strongly disagree. The lowest rating value one indicates strongly disagree and the highest value indicates strongly agree.

(Refer Slide Time: 20:25)

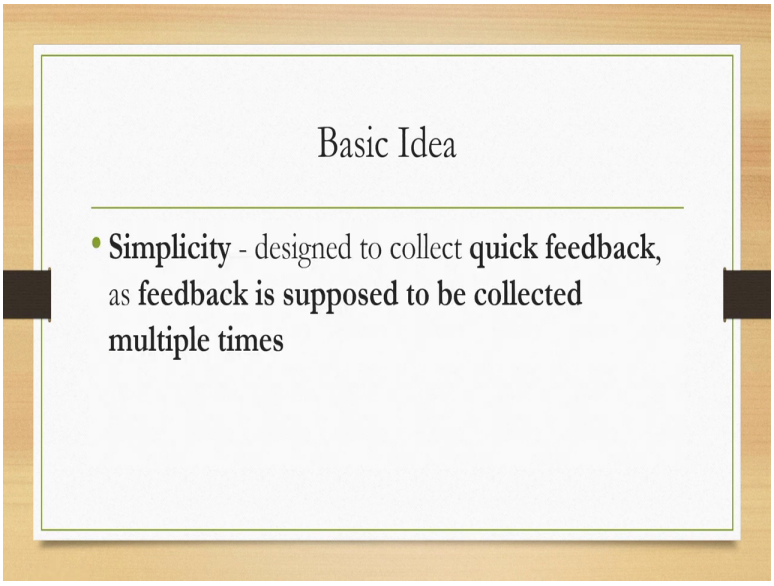


So, what are the questions. So, there are three items or statements in the questionnaire item number one I am satisfied with the ease of completing the tasks in this scenario. So, a task is performed by the user following a task scenario and user has to provide rating for this statement or item which is I am satisfied with the ease of completing the tasks in this scenario. Then second one is I am satisfied with the amount of time it took to complete the tasks in this scenario.

Again here the user is expected to provide a rating in the seven point rating scale for this item after the task is completed. Then the third item in the questionnaire is I am satisfied with the support information which includes online help messages documentation when completing the tasks. So, this item also needs to be rated in that 7 point rating scale by the user after the task is complete. So, the user is given a task and then asked to provide ratings for all these questions or items these are basically statements rather than questions.

So, it is better to call them items but the whole set of three items we can call questionnaire using the standard terminology. Just to repeat statement one or item one is I am satisfied with the ease of completing the tasks in this scenario. Then item number two I am satisfied with the amount of time it took to complete the tasks in this scenario. And item number three I am satisfied with the support information when completing the tasks where the support information include online help messages and documentation.

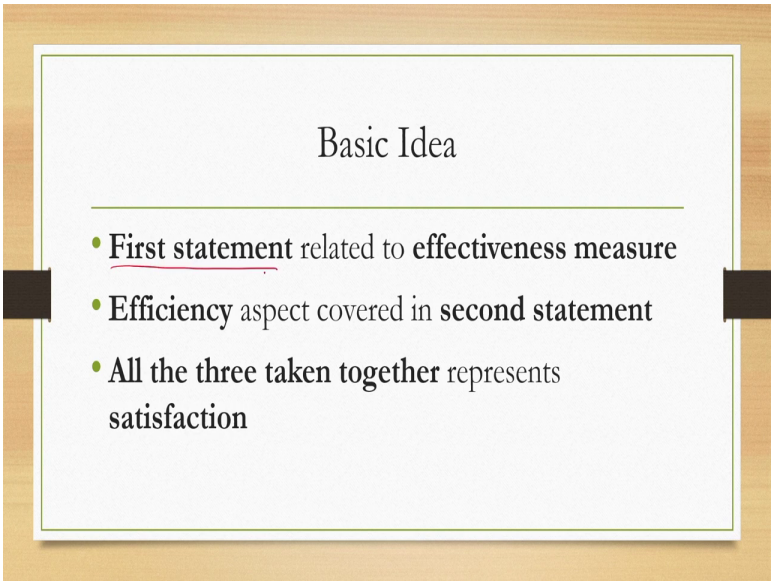
(Refer Slide Time: 22:30)



As you can see from the questionnaire or rather the items in the questionnaire it is simple so the questionnaire is designed to collect quick feedback and this is suitable because we are supposed to collect feedback multiple times in the early phase of design. So, this questionnaire along with this method supports that idea. Now let us try to understand how the feedback provided by rating on these three items can be related to the idea of usability and usability issues.

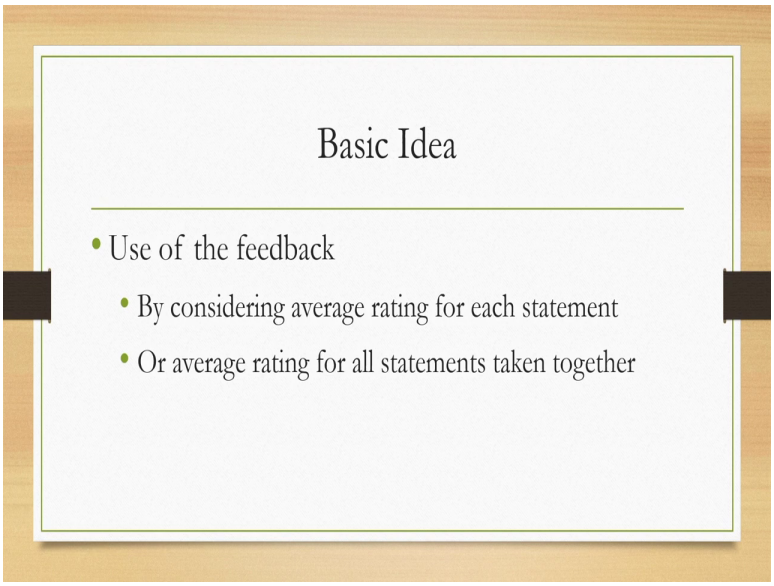
So, when we discussed usability if you may recollect we mentioned three things as per ISO standard definition of usability these are effectiveness, efficiency and satisfaction.

(Refer Slide Time: 23:28)



First statement in the ASQ questionnaire is related to the effectiveness measure that we have mentioned in the definition of usability. Efficiency aspect is covered in the second statement if you look at the statements closely and all the three statements taken together represents satisfaction. So, the ratings for all the three statements together represent satisfaction whereas fast statement refers to effectiveness and second one is refers to efficiency.

(Refer Slide Time: 24:01)



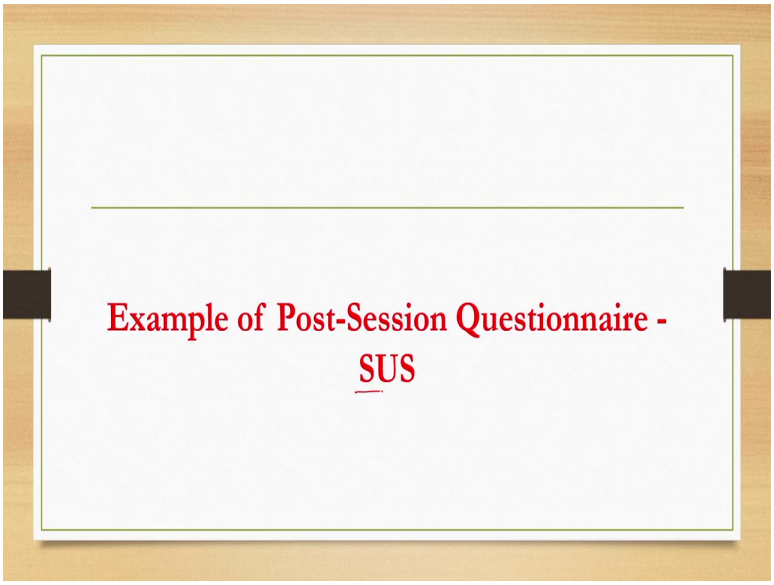
So, how we can make use of the feedback or the ratings that we get from the users. So, we can actually take an average rating for each statement. Suppose we are collecting the rating from five end users after they were asked to perform some tasks using our proposed design of the interface.

Now to perform the tasks it is not necessary to implement the interface rather some prototypes are fine and the task scenario needs to be mentioned the way we did it in cognitive walkthrough.

But now here instead of expert evaluators we have end users to evaluate the interfaces. So, now after each task is performed we can use the questionnaire to get the ratings and once we collect ratings from all the five users we can take average rating for each item or we can take average rating for all the statements taken together either of these are fine. And based on the average rating we can take a decision.

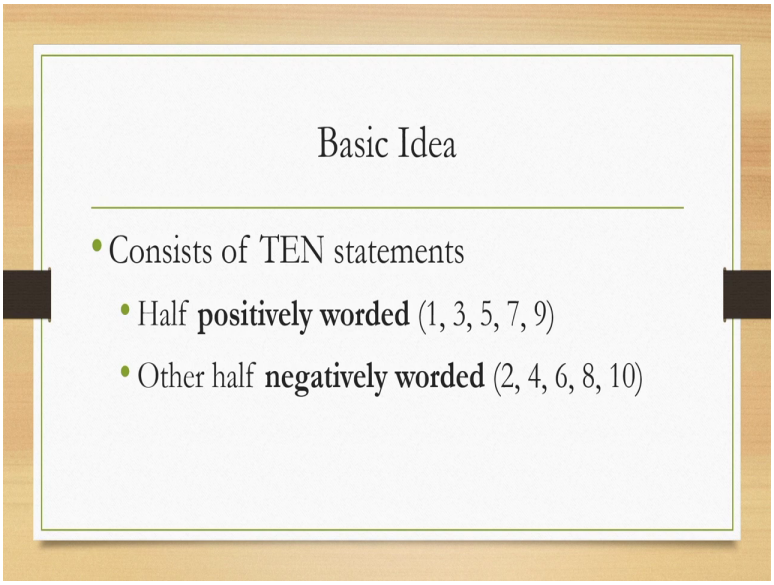
For example suppose the rating is closer to one that means users are mostly disagree with the statements in that case there are likely to be usability issues and we need to revise the design. If the average ratings are closer to 7 which is the highest rating then probably issues are less and no need to refine the design further that can be one of the way to conclude from the data that we collect. So, that is about post task questionnaire.

(Refer Slide Time: 25:56)



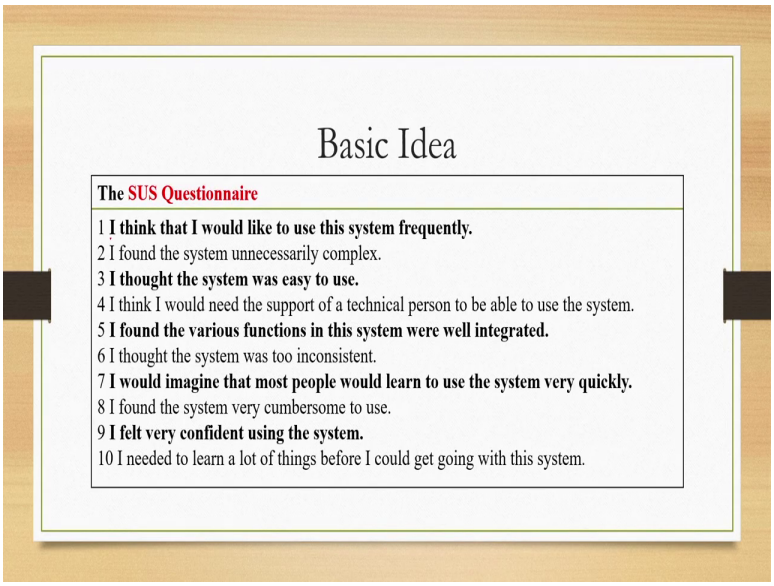
Now let us try to understand post session feedback collection using questionnaire with respect to one example here we will be using the example of system usability scale or SUS questionnaire.

(Refer Slide Time: 26:02)



Now SUS questionnaire consists of 10 statements. Again here we are using the term statement rather than questions because these are in the form of statements. One interesting thing is that half of these statements are positively worded and the other half are negatively worded. So, statement number 1, 3, 5, 7 and 9 odd number statements are positively voted and even number statements 2, 4, 6, 8 and 10 are negatively worded. Let us have a look at these statements.

(Refer Slide Time: 26:41)



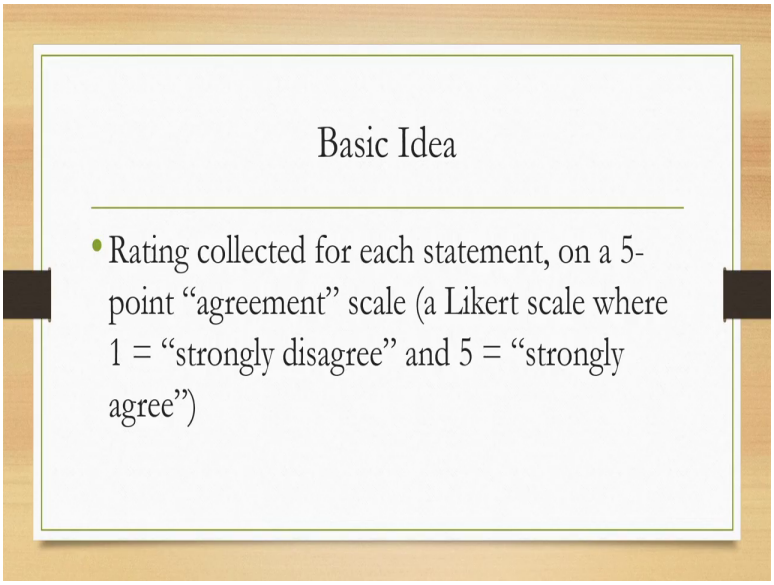
Statement number one I think that I would like to use this system frequently. As you can see this is positively worded statement. So, one is positively worded. Statement number two says I found the system unnecessarily complex again this is a negatively worded statement. Statement number

three I thought the system was easy to use another example of a positively worded statement. Statement 4 I think I would need the support of a technical person to be able to use the system this is a negatively worded statement.

Statement 5 I found the various functions in this system were well integrated as you can see this is another positively worded statement. Six says I thought the system was too inconsistent example of negatively worded statement. Seven I would imagine that most people would learn to use the system very quickly a positively worded statement. Eight is I found the system very cumbersome to use a negatively worded statement.

Statement nine I felt very confident using the system again a positively added statement whereas statement 10 says I needed to learn a lot of things before I could get going with this system which is a negatively worded statement. So, here as we just have seen the odd number statements namely 1 3 5 7 9 are positively worded whereas even number statements 2 4 6 8 and 10 are negatively worded together they constitute the SUS questionnaire having these 10 items.

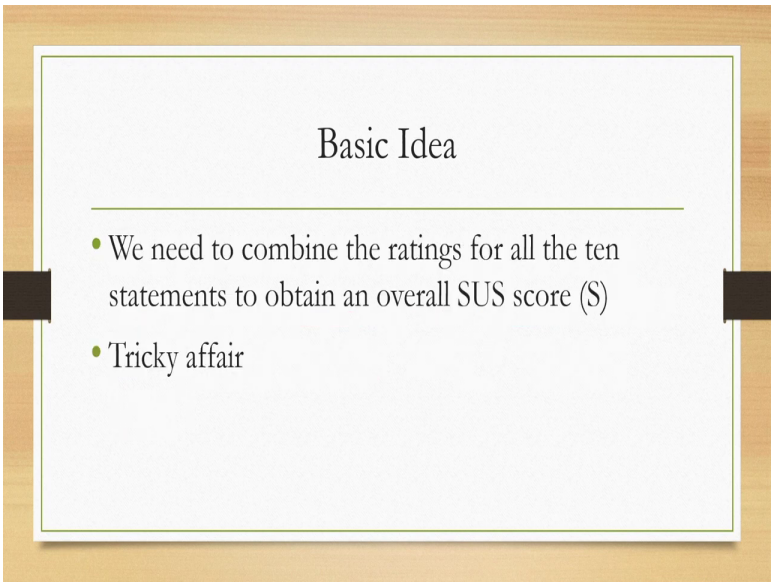
(Refer Slide Time: 29:05)



So, what to do with these items? So, post session the questionnaire is given to the users and they are asked to provide rating for each statement on a five point agreement scale which is a Likart scale where one indicates strongly disagree and five indicates strongly agree. So, for each item

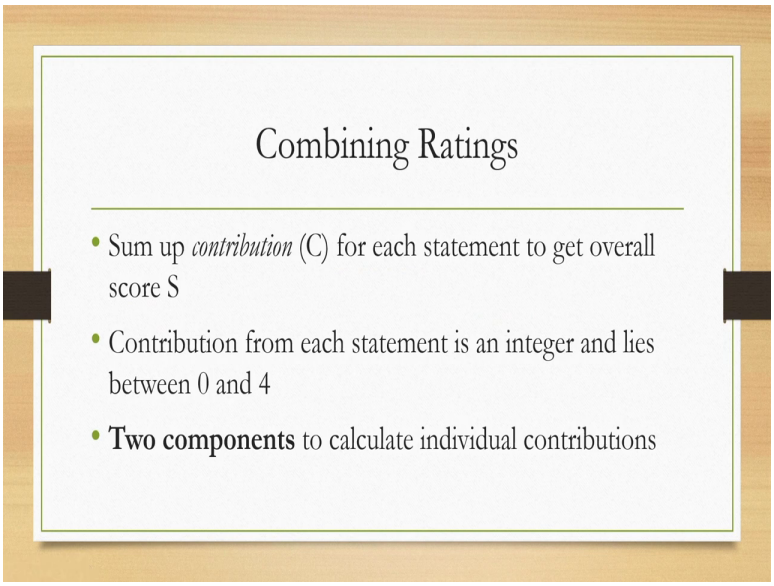
the user has to provide feedback in the form of ratings in a five point agreement scale where 1 indicates strongly disagree and 5 indicates strongly agree.

(Refer Slide Time: 29:45)



Unlike in the case of ASQ here combining these ratings is slightly different. So, we need to combine the ratings for all the 10 statements to obtain an overall SUS score which is a tricky appear it is not as simple as simply taking an average which we did in the case of the earlier one.

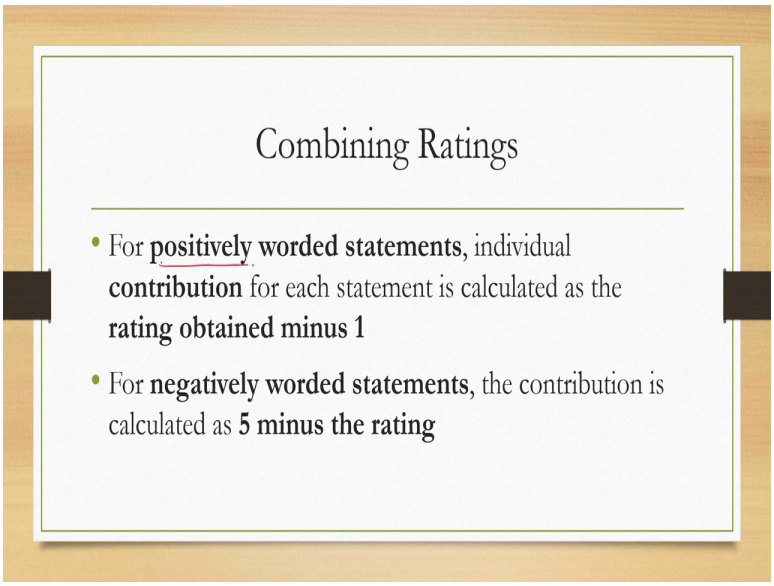
(Refer Slide Time: 30:13)



So, how do we perform the combination? We first sum up the contribution let us call it C for each statement to get overall score S. Now this contribution from each statement is an integer

which lies between the value 0 and 4, two components are there to calculate individual statement contributions.

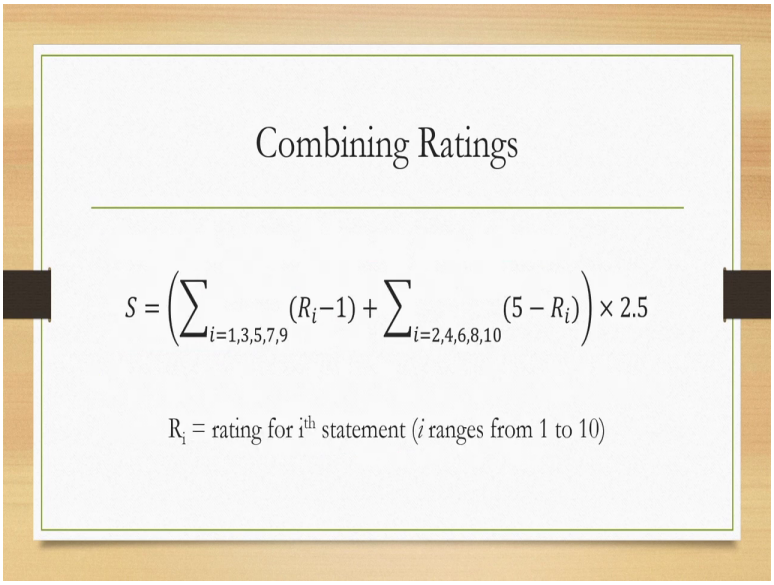
(Refer Slide Time: 30:50)



For positively worded statements that is statements which are odd numbered the individual contribution for each statement is calculated as the rating obtained minus 1. So, it is R minus 1. So, if the user provides some rating for that statement we deduct 1 from it to get the individual contribution. For negatively worded statements the contribution is calculated as 5 minus the rating. So, here it is 5 minus R.

So, whatever ratings the users provided for the negatively worded items or statements that is the items which are even numbered we basically calculate the individual contribution for that particular statement by subtracting the rating from 5.

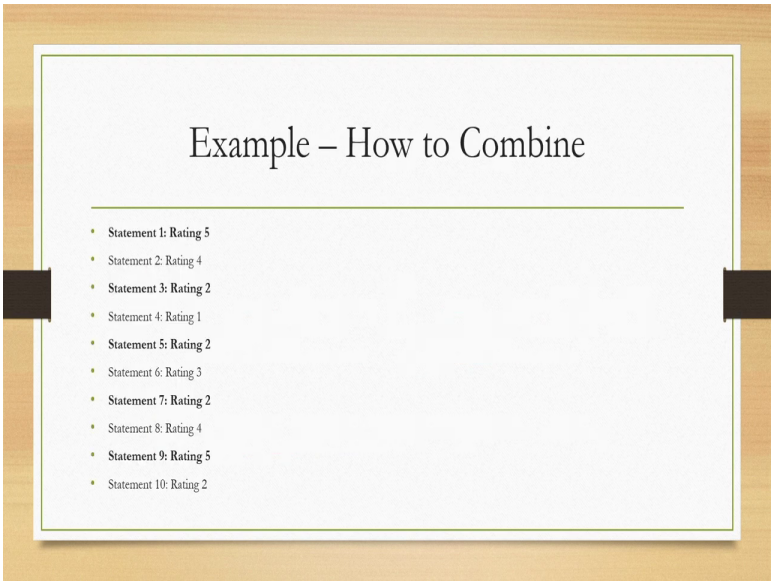
(Refer Slide Time: 31:47)



And then we sum up these contributions as mentioned earlier. So, total SUV score is calculated as summation for individual contributions by all the positively worded statements that is even number statements R i - i where i varies from 1 3 5 7 9 within this set and contribution for the statements which are negatively worded 5 minus R i where i varies between these numbers 2 4 6 8 and 10. Now we sum it up together all the 10 contributions and then multiply with a factor 2.5.

R i here indicates the rating for the ith statement and individual cases accordingly we vary i. So, after multiplication the number that we get is the SUS score.

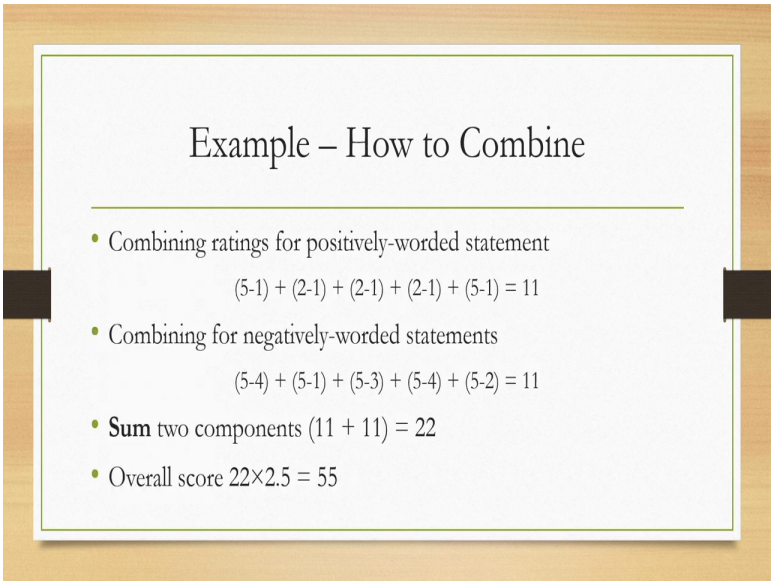
(Refer Slide Time: 32:54)



Let us try to understand it with an example. Suppose for statement one in a given situation where the user was asked to use some particular interface design and provide the ratings the users have provided the following ratings. For statement 1 rating is 5 for statement 2 rating is 4 for statement 3 rating is 2 statement 4 rating one statement 5 rating 2 statement 6 rating 3 statement 7 again rating 2 statement 8 rating 4statement 9 rating 5 and statement 10 rating 2.

Now these ratings are provided by an user for a given design on the basis of the SUS questionnaires. So, we now need to combine these ratings to get the SUS score for that particular user.

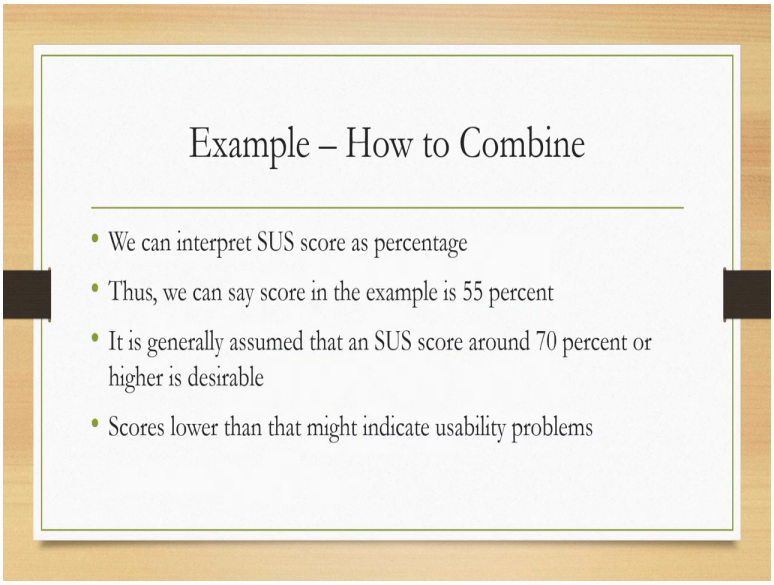
(Refer Slide Time: 33:59)



So, there are two components we first have to combine ratings for positively worded statements to get the contributions from those statements. So, it will be five minus one plus two minus one plus two minus one plus two minus one plus five minus one or eleven. Then we have to combine for negatively varied statements five minus rating that is five minus four five minus one five minus three five minus four and five minus two this also and then we add them together.

So, this also results in the number eleven then we sum these two components. So, we get eleven plus eleven that is 22 and then we get the overall score by multiplying the sum with the factor 2.5 to get 55.

(Refer Slide Time: 34:55)



So, how to interpret this number so, we got a score 55. So, what it means we can interpret this core as a percentage. So, we can say that the score in the example is 55%. So, we got the number 55. So, we can say that SUS score is 55%. It is generally assumed based on several studies that an SUS score around 70% or more is desirable for a system or an interface design to be considered not having significant usability issues.

If SUS score is found out to be less than 70% then there are likely to be some usability issues and it is better to redesign the interface. So, scores lower than the 70 percentage point might indicate usability problems and it is preferable to revise or refine such designs. So, that is how we can make use of the SUS score. So, that is what we can do with user based quick evaluation. Here we have seen two methods.

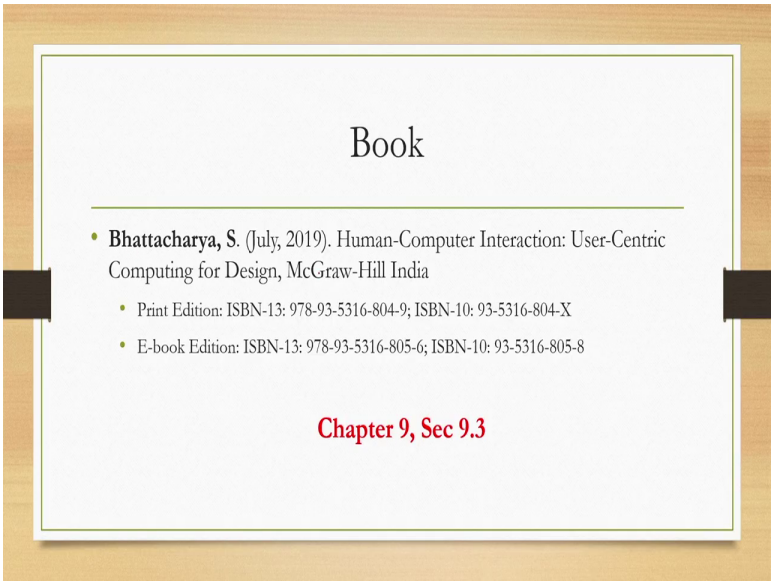
So, the basic setup remains the same as in the case of expert evaluation. So, we have some tasks and prototypes to carry out the tasks preferably vertical prototype here. So, that tasks can be carried out then what we can do is we can collect feedback using rating scales from the users after they were asked to perform those tasks in a controlled setting using the prototypes. So, unlike in the expert evaluation method here the feedback is collected using rating skills.

So, we have two types of questionnaires to collect feedbacks from two points one is after each task is carried out and second is after all the tasks are carried out that means at the end of the

session. ASQ is one example questionnaire that we have discussed to collect feedback after each task and SUS is one example questionnaire that we have discussed to collect feedback after the whole session. As we have discussed each of these methods has its own positive and negative sides.

Post task questionnaire is likely to reveal more about the specific components of the interface that has been used to carry out the task. Whereas post session questionnaire is likely to reveal more about the general perception of the user about the whole system rather than task specific perception it is preferable to use both as much as possible. So, that is what we can do to quickly evaluate prototype with end users note that the requirement that the evaluation takes place in a rapid manner is very important and these self reporting methods can satisfy that requirement with that we end the topic here.

(Refer Slide Time: 38:26)



So, whatever I have discussed can be found in this book human computer interaction you are requested to refer to chapter 9 section 9.3 to know in details about these methods. I hope you have enjoyed the content and I am looking forward to meet you all in the next lecture, thank you and goodbye.