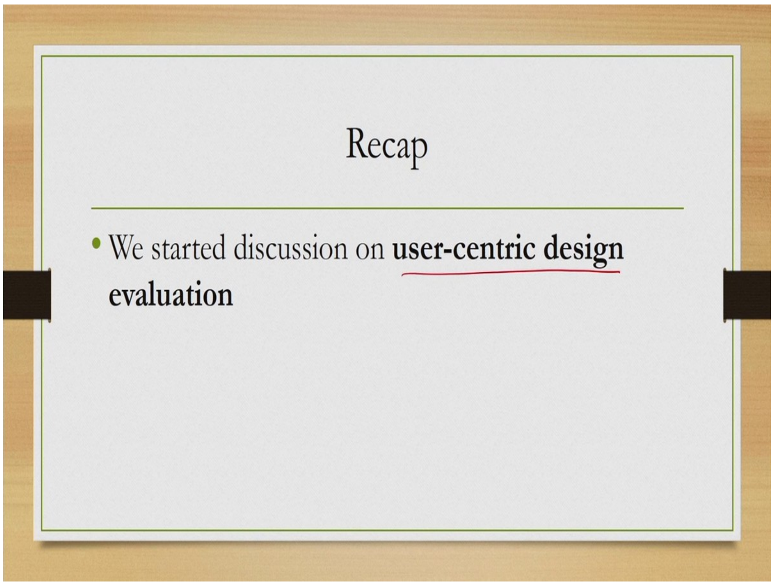
User-Centric Computing for Human-Computer Interaction

Prof. Samit Bhattacharya
Department of Computer Science and Engineering
Indian Institute of Technology, Guwahati

Lecture – 30
User Evaluation, Empirical and Model-Based Evaluation

Hello and welcome to lecture number 30, in the course User-Centric Computing for Human Computer Interaction.

(Refer Slide Time: 00:46)

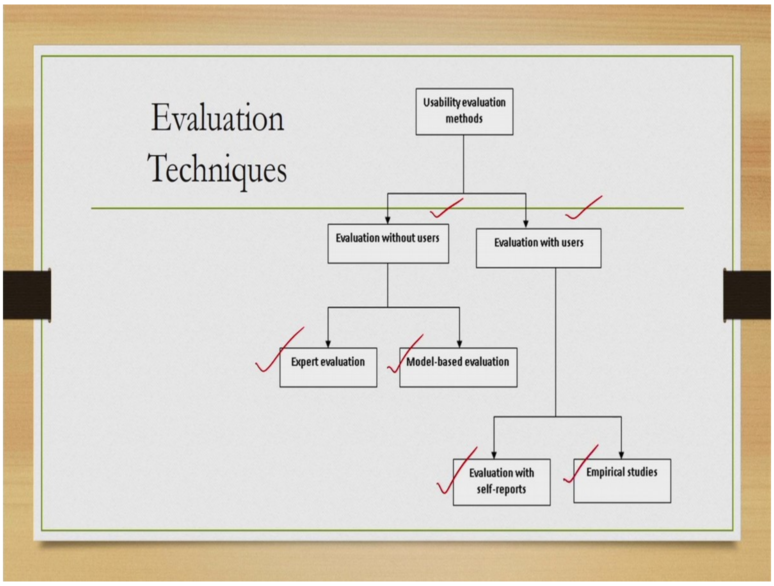


Recap

- We started discussion on **user-centric design evaluation**

So, what we have discussed in the previous lecture? We started our discussion on how to evaluate user-centric designs. And, there we mentioned that there are many evaluation techniques.

(Refer Slide Time: 01:02)

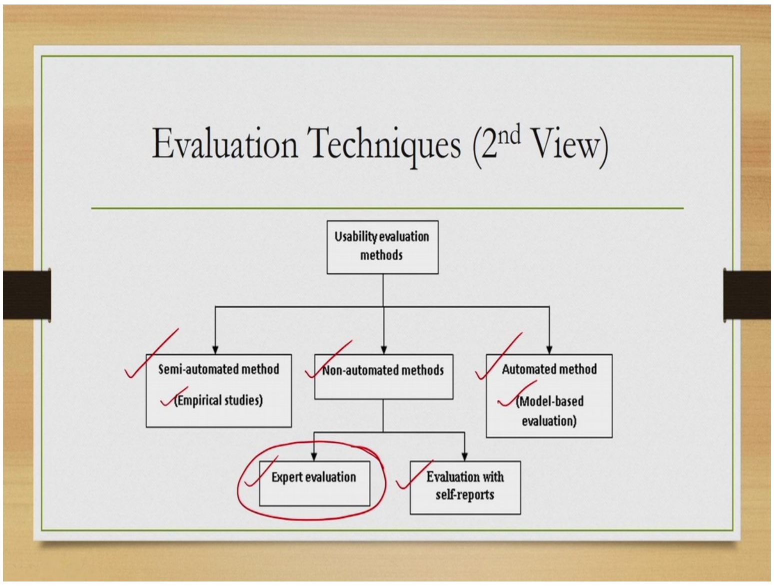


Evaluation Techniques

There are different ways to view those techniques. One way is to view the techniques in the form of this hierarchy where we have two broad categories evaluation without users and evaluation with users. Under evaluation without users, we have expert evaluation techniques and model based evaluation techniques. Under evaluation with users, we have evaluation with self-reports and empirical studies.

There is another way to view it.

(Refer Slide Time: 01:35)

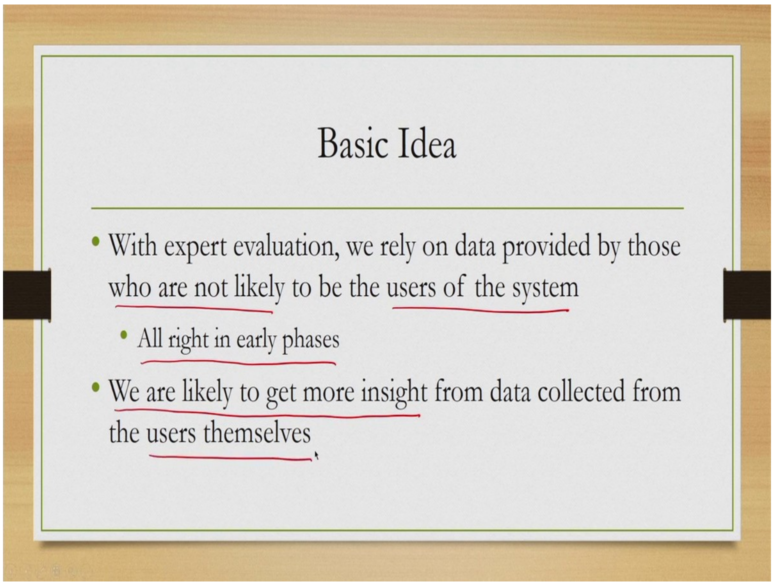


Evaluation Techniques (2nd View)

So, we broadly have three categories semi-automated methods, non automated methods and automated methods. These terms indicate the methods that we apply to analyze the data that we collect during evaluation.

Now, under semi-automated method, we have empirical studies; under automated method, we have model based evaluations and under non automated methods, we have expert evaluations and evaluations with self-reports. Among them, we have already discussed in the previous lecture expert evaluations. So, today we are going to discuss about the remaining methods.

(Refer Slide Time: 02:31)



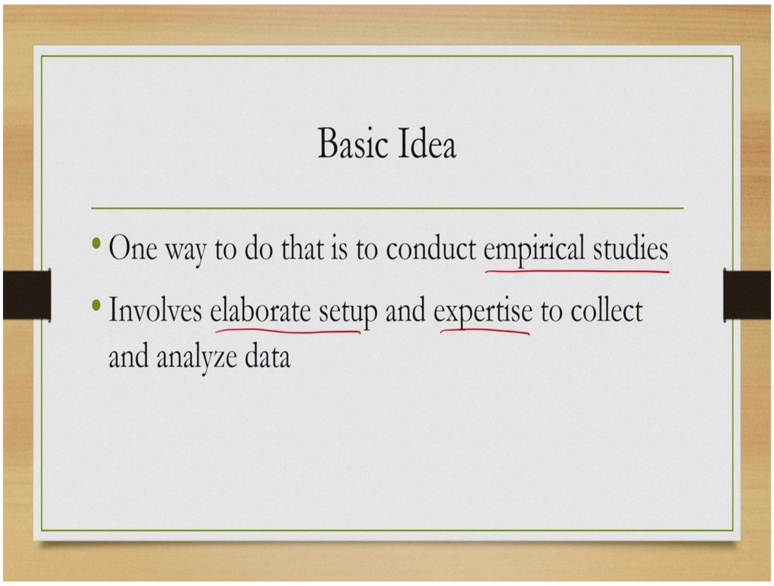
Basic Idea

- With expert evaluation, we rely on data provided by those who are not likely to be the users of the system
  - All right in early phases
- We are likely to get more insight from data collected from the users themselves

Let us start with user evaluation. So, with expert evaluation, we rely on data that is provided by those who are not likely to be the users of the system. For example, the designers or the other skill designers; if you include some end users of course, that is a separate thing, but mostly when we perform expert evaluation. Typically, those were not likely to be the end users are involved although they are skilled in their field.

However, it is all right in the early phases when we need to go for rapid prototyping and evaluations and there are many iterations. But of course, this is a fact that we are likely to get more insight, if we collect the data from the users themselves; that is obvious.

(Refer Slide Time: 03:33)

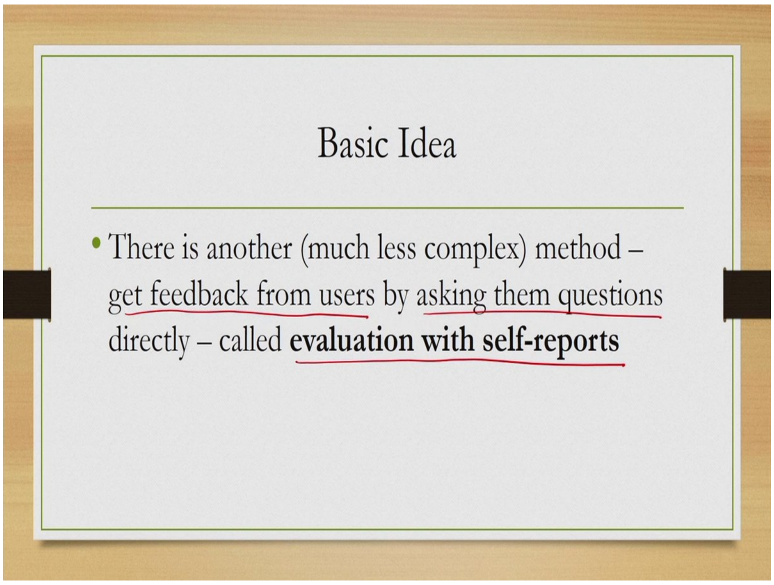


Basic Idea

- One way to do that is to conduct empirical studies
- Involves elaborate setup and expertise to collect and analyze data

So, how we can collect data from the user? One way is to go for empirical studies, we have already discussed it before. What it requires is elaborate setup and expertise to collect and analyze the data.

(Refer Slide Time: 03:57)



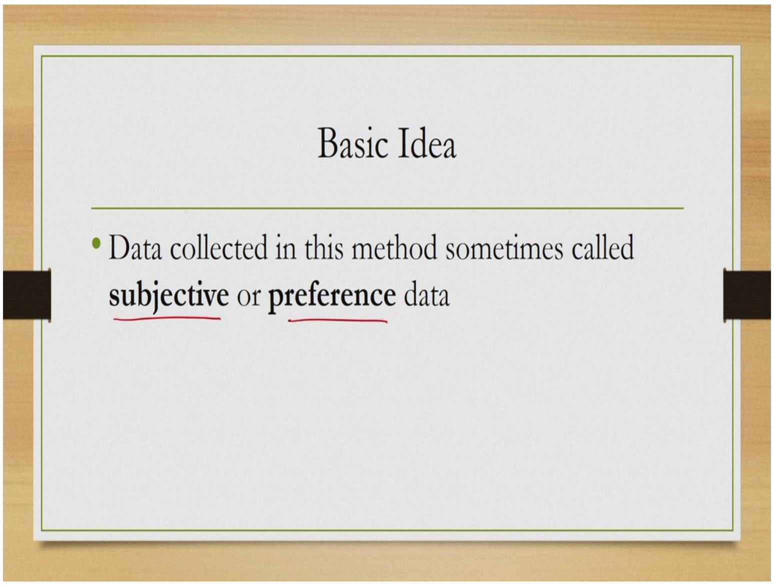
Basic Idea

- There is another (much less complex) method – get feedback from users by asking them questions directly – called **evaluation with self-reports**

There is another way which is much less complex, which is directly get feedback from the users by asking them questions.

Now, when we are going for this approach, where we are asking the users questions and taking their feedback on those questions rather than asking them to perform some tasks will log their data, then analyze it as we do in empirical study. That approach is known as evaluation with self-reports. So, how we can do this?

(Refer Slide Time: 04:40)

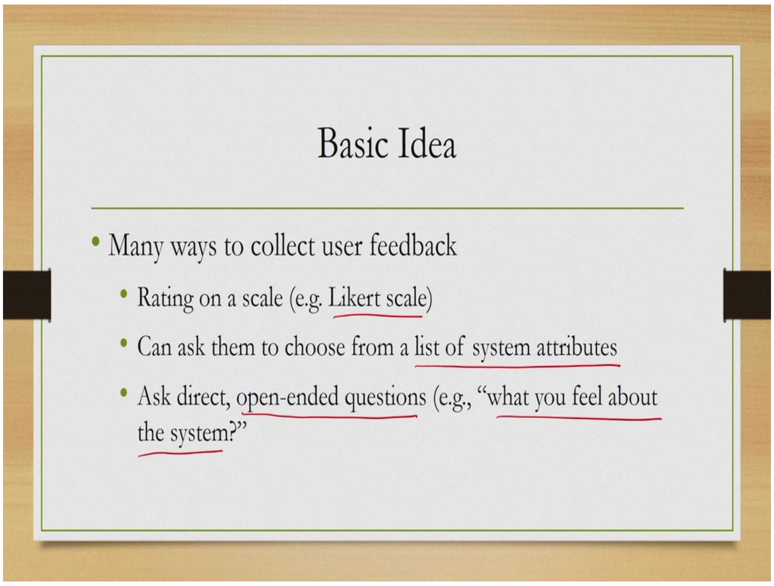


Basic Idea

- Data collected in this method sometimes called **subjective** or **preference** data

The data that we collect in this method sometimes called subjective or preference data. The question is how we can do this.

(Refer Slide Time: 04:56)

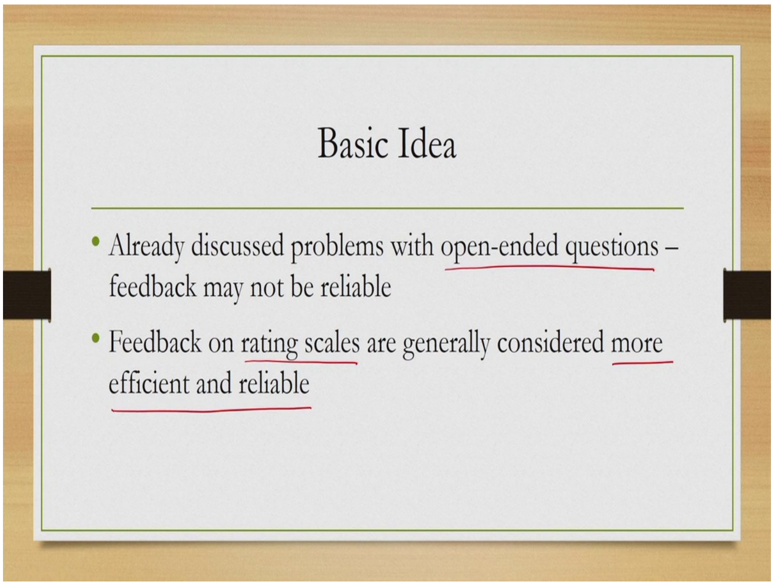


Basic Idea

- Many ways to collect user feedback
  - Rating on a scale (e.g. Likert scale)
  - Can ask them to choose from a list of system attributes
  - Ask direct, open-ended questions (e.g., "what you feel about the system?"

Let us try to see. So, there are many ways. One is we can ask for their rating on a scale such as a Likert scale which we have already discussed before, then we have a design and we can ask them to choose from a list of system attributes that they want to have in the system or we can ask open ended questions such as, “what you feel about the system” or how we can make the system better.

(Refer Slide Time: 05:33)



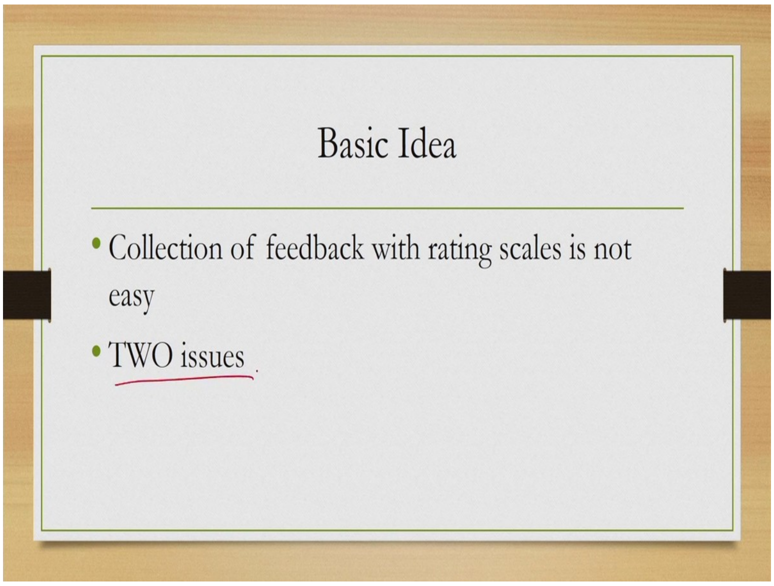
Basic Idea

- Already discussed problems with open-ended questions – feedback may not be reliable
- Feedback on rating scales are generally considered more efficient and reliable

Earlier, we have discussed the problem with open-ended questions. There is no guarantee that the feedback that we get is reliable. On the other hand, feedback we received through some rating scales, we can generally consider such feedback to be more efficient and reliable.

So, when we are going for feedback and if we are using a rating scale to get the ratings as feedback, this is considered to be more efficient and reliable than going for feedback through open ended questions.

(Refer Slide Time: 06:19)

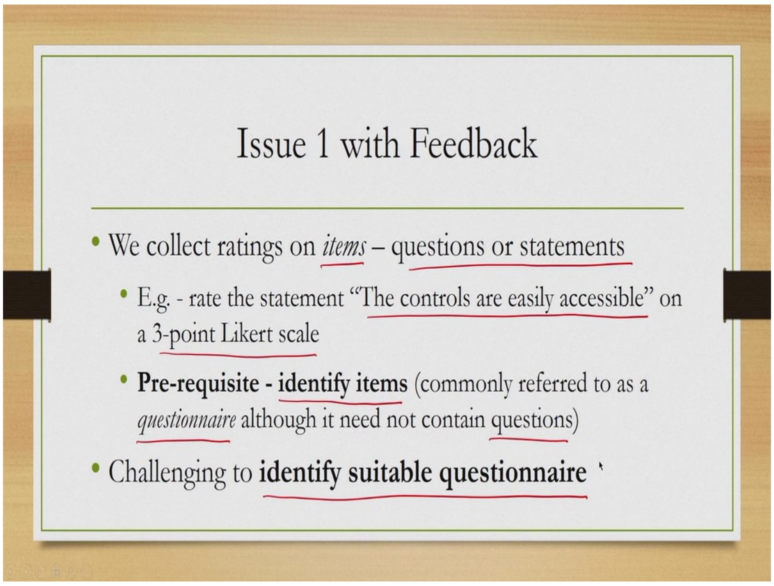


Basic Idea

- Collection of feedback with rating scales is not easy
- TWO issues

But, the problem is collection of feedback with rating scale is also not easy and there are broadly two issues. What are these issues?

(Refer Slide Time: 06:34)



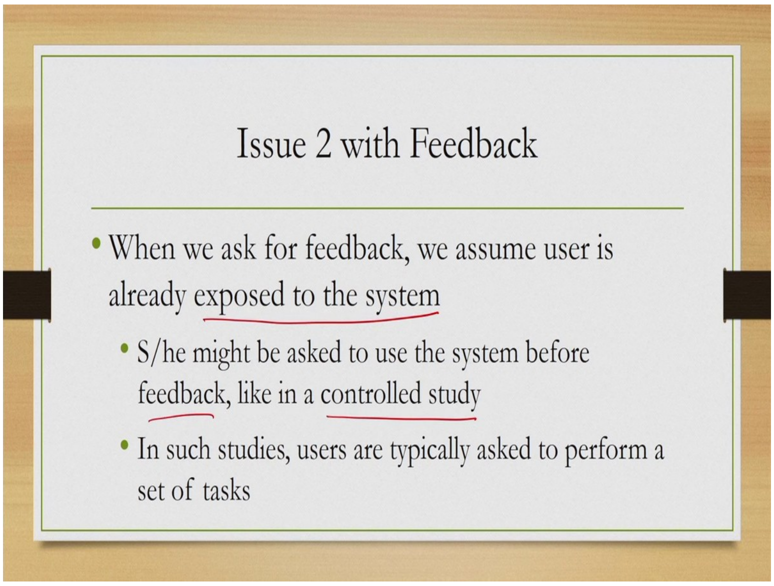
Issue 1 with Feedback

- We collect ratings on *items* – questions or statements
  - E.g. - rate the statement "The controls are easily accessible" on a 3-point Likert scale
  - **Pre-requisite - identify items** (commonly referred to as a *questionnaire* although it need not contain questions)
- Challenging to **identify suitable questionnaire**

First of all when we say that we want to collect ratings, what we mean is that we are collecting ratings on some items, which are some questions or statements. For example, we may like to collect rating on a 3-point Likert scale for the statement "the controls are easily accessible".

So, the prerequisite for going for rating best feedback is that, we need to identify these items or questions of statements. So, they are most often known as a questionnaire although, it is not necessary that the items refer to any questions. It may be some statements also. And, as it is obvious, it is always a challenging task to identify suitable questionnaire for ratings. There is another issue.

(Refer Slide Time: 07:33)

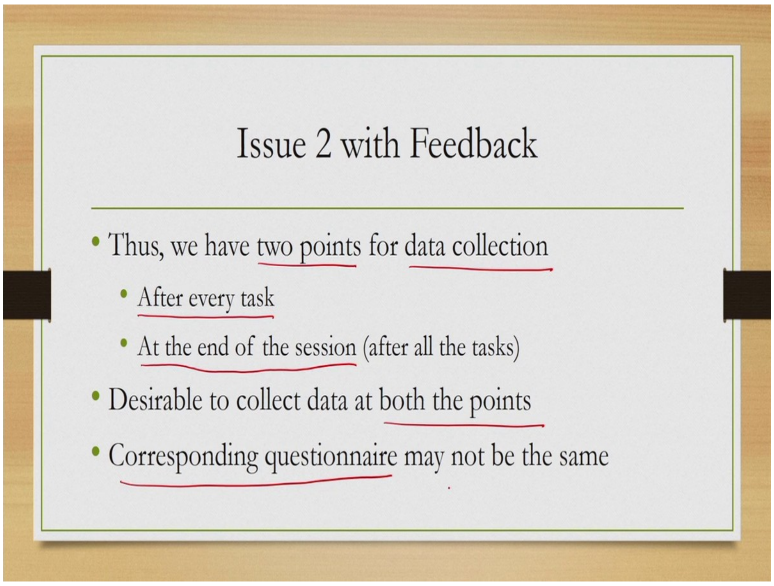


Issue 2 with Feedback

- When we ask for feedback, we assume user is already exposed to the system
  - S/he might be asked to use the system before feedback, like in a controlled study
  - In such studies, users are typically asked to perform a set of tasks

So, when we ask for feedback, we assume that the user is already exposed to the system. And, how we can expose the user to the system? Let us consider a control study where we ask the user to use the system before feedback. And, use the system means they are asked to perform some tasks.

(Refer Slide Time: 08:03)



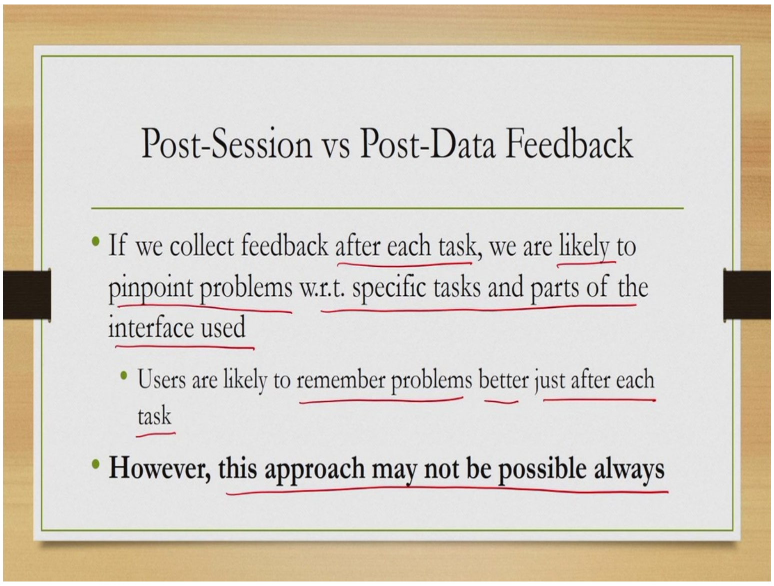
Issue 2 with Feedback

- Thus, we have two points for data collection
  - After every task
  - At the end of the session (after all the tasks)
- Desirable to collect data at both the points
- Corresponding questionnaire may not be the same

Now, when they are performing tasks and when they finish performing the tasks, these two are different points of interest. So, after every task we can take their feedback or at the end of the completion of all the tasks or at the end of the session, we can collect the feedback. So, there are two points for data collection. It is always desirable to collect data at both the points; at the end of every task and at the end of the session.

However, that may not be possible always, but if that is possible then the corresponding questionnaire need not be the same, they may be different. So, the questionnaire that we used to collect data at the end of every task and the questionnaire that we used to collect data at the end of the session may be different. So, that is challenging to come up with two sets of questionnaires to collect data different points.

(Refer Slide Time: 09:12)



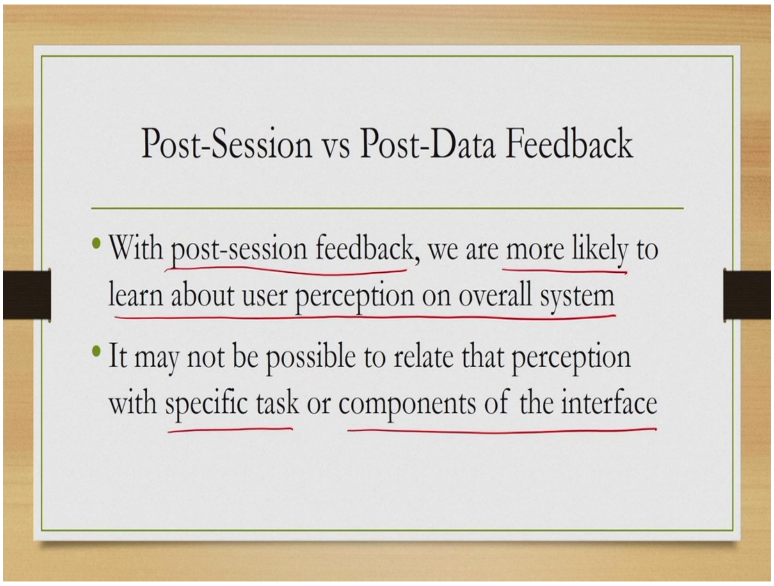
Post-Session vs Post-Data Feedback

- If we collect feedback after each task, we are likely to pinpoint problems w.r.t. specific tasks and parts of the interface used
  - Users are likely to remember problems better just after each task
- **However, this approach may not be possible always**

There are advantages and disadvantages for both the points of data collections. So, if we collect feedback after each task, then we are likely to pinpoint some problems with respect to specific tasks and parts of the interface used. Because, the users are likely to remember problems related to the tasks and the parts of the interface that they have used to carry out the task better just after each task.

So, if you ask them to pinpoint problems related to tasks or parts of the interfaces after a long time, they may have forgotten many things and may not be able to give you proper rating as feedback. But, as I said this approach may not be possible always; if, there is a large number of tasks given and the users time is very precious. So, after every task if we ask the user to give feedback, then the entire process may take long time which the user may not be able to afford.

(Refer Slide Time: 10:31)



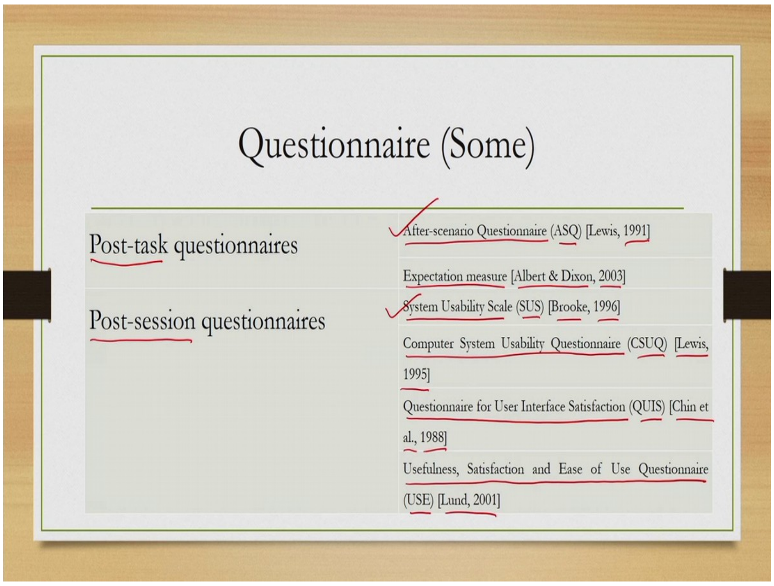
Post-Session vs Post-Data Feedback

- With post-session feedback, we are more likely to learn about user perception on overall system
- It may not be possible to relate that perception with specific task or components of the interface

On the other hand the post session feedback is more common. So, at the end of all the tasks we typically ask for feedback. But if we go for that, then we are more likely to learn about user perception on overall system rather than, about specific tasks or portion of the interfaces.

Now, that overall perception may not be relatable to any specific task or components of the interface. So, we may be interested to find out a specific task or specific components that may create problem. But from the overall impression, it may not be possible to identify specific things in the design.

(Refer Slide Time: 11:32)



Questionnaire (Some)

Post-task questionnaires	After-scenario Questionnaire (ASQ) [Lewis, 1991] Expectation measure [Albert & Dixon, 2003]
Post-session questionnaires	System Usability Scale (SUS) [Brooke, 1996] Computer System Usability Questionnaire (CSUQ) [Lewis, 1995] Questionnaire for User Interface Satisfaction (QUIS) [Chin et al., 1988] Usefulness, Satisfaction and Ease of Use Questionnaire (USE) [Lund, 2001]

Having said that let us now try to see the issues concerning design of questionnaires. So, as I said identifying a suitable questionnaire is challenging, fortunately there are already many standard questionnaires available. And, we can make use of those rather than trying to design it our self. Now, these questionnaires available for both the points of data collection namely post task and post session.

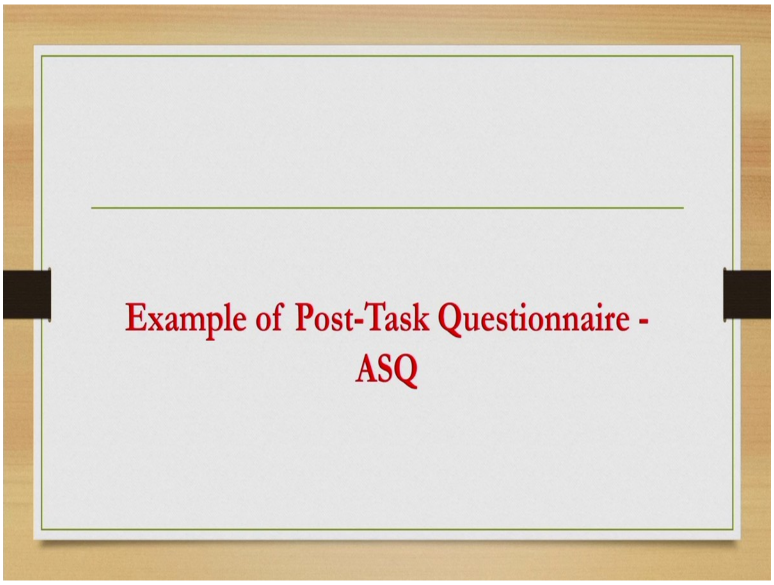
So, for post task there are two popular questionnaires available, After-scenario questionnaire or ASQ, it was proposed by Lewis in 1991 and expectation measure proposed by Albert and Dixon in 2003. There may be many other questionnaires available, but these two are well known.

Similarly, for post session data collection we have many questionnaires available, the more popular ones are System Usability Scale or SUS proposed by Brooke in 1996, Computer System Usability Questionnaire CSUQ proposed by Lewis in 1995.

Questionnaire for User Interface Satisfaction or QUIS, QUIS proposed by Chin et al in 1988 and Usefulness, Satisfaction and Ease of Use Questionnaire or USE proposed by Lund in 2001. Again these are not the only questionnaires available for collecting feedback at the end of the session, there are many other questionnaires, but these are popularly used and well known.

In order to get some understanding of this questionnaires, we will go through a couple of them one for each category. So, we will discuss the ASQ questionnaire and the SUS questionnaire to give you an understanding of how the questionnaires are designed and how they are used.

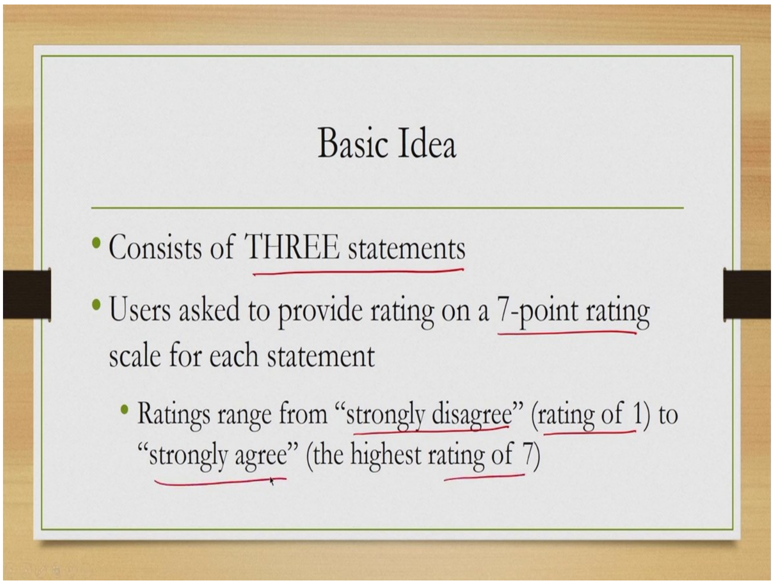
(Refer Slide Time: 14:00)



Example of Post-Task Questionnaire - ASQ

Let us start with the post task questionnaire ASQ.

(Refer Slide Time: 14:07)

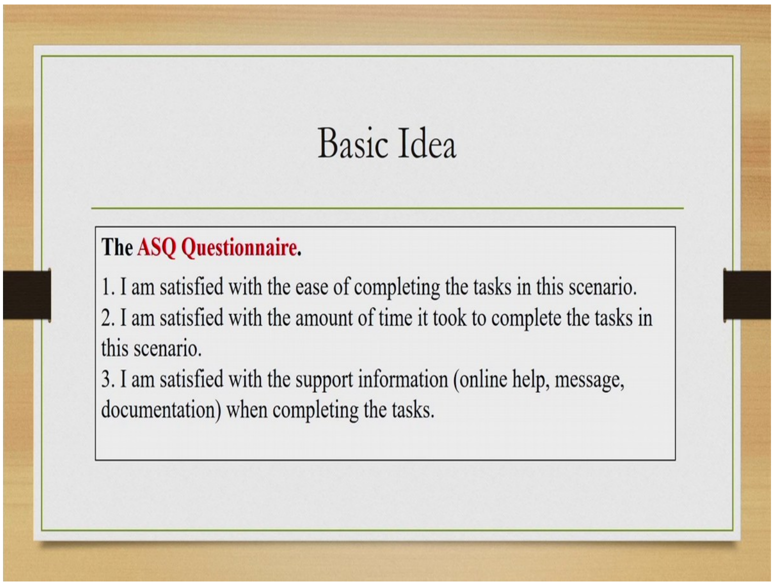


Basic Idea

- Consists of THREE statements
- Users asked to provide rating on a 7-point rating scale for each statement
  - Ratings range from "strongly disagree" (rating of 1) to "strongly agree" (the highest rating of 7)

Now, this questionnaire consists of three statements. And, each user is asked to provide rating on a 7-point rating scale for each statement. So, in this rating scale rating 1 indicates “strongly disagree” and rating 7 indicates “strongly agree”. What are these statements?

(Refer Slide Time: 14:40)



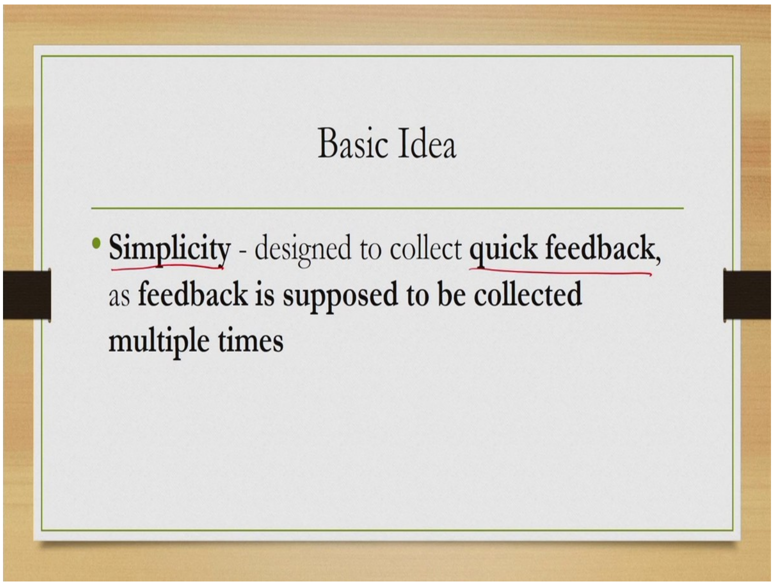
Basic Idea

The ASQ Questionnaire.

1. I am satisfied with the ease of completing the tasks in this scenario.
2. I am satisfied with the amount of time it took to complete the tasks in this scenario.
3. I am satisfied with the support information (online help, message, documentation) when completing the tasks.

The first statement is I am satisfied with the ease of completing the tasks in this scenario. Second statement is I am satisfied with the amount of time it took to complete the tasks in this scenario. And, the third statement is I am satisfied with the support information online help, message, documentation when completing the tasks. These are the three statements, which form the questionnaire ASQ questionnaire, for each statement at the end of each task the user is asked to give a rating in a 7 point rating scale where 1 indicates strongly disagree and 7 indicates strongly agree.

(Refer Slide Time: 15:36)

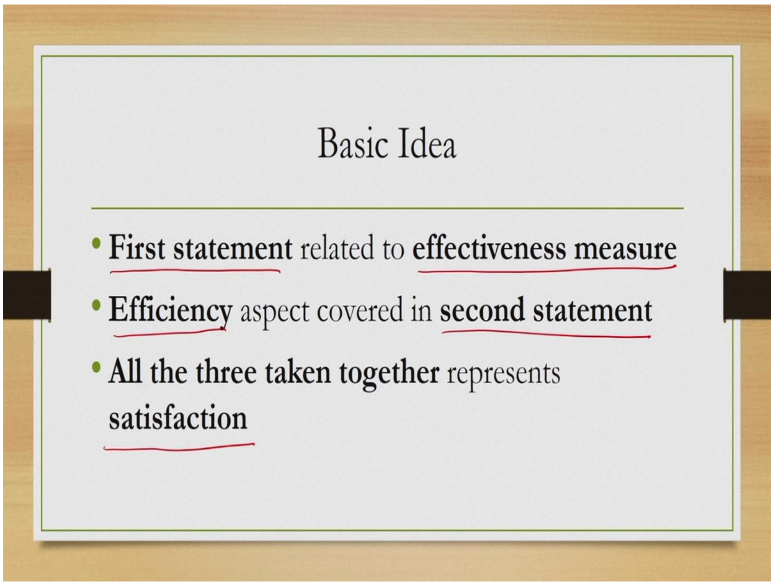


Basic Idea

- Simplicity - designed to collect quick feedback, as feedback is supposed to be collected multiple times

One thing you may note here is that the questions are very simple. It is so, since it is designed to collect quick feedback because the feedback is supposed to be collected at the end of each task and therefore, it has to be collected many times. So, to make it quick very simple statements have been used to from the questionnaire.

(Refer Slide Time: 16:09)



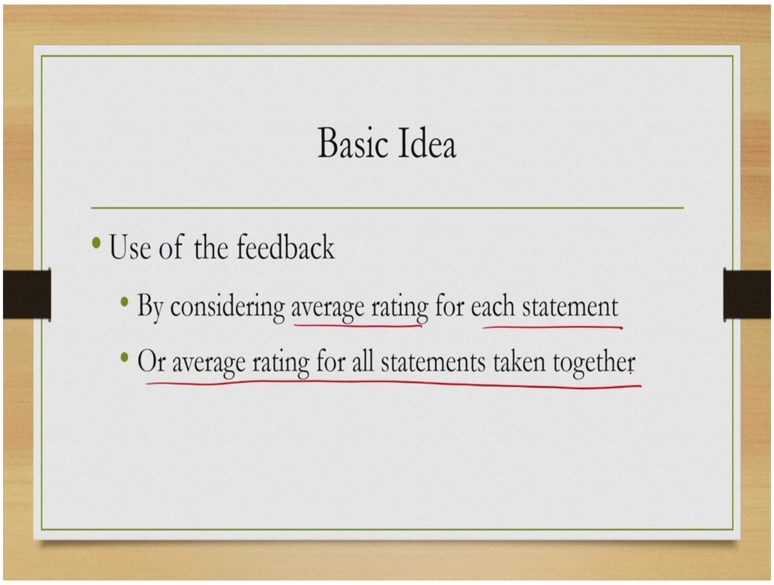
Basic Idea

- First statement related to effectiveness measure
- Efficiency aspect covered in second statement
- All the three taken together represents satisfaction

Now, once we collect rating what to do? There are two ways, but before that let us see how they are related to usability. The first statement can be related to the effectiveness measure, effectiveness measure of usability. The second statement can be related to the efficiency measure of usability. And, all the three taken together can be used to represent satisfaction.

So, these three as you may recollect a part of the definition of ISO standard. So, these three measures effectiveness efficiency and satisfaction refers to the ISO standard definition of usability and this ASQ questionnaire is designed to help identify or help to evaluate these components of usability.

(Refer Slide Time: 17:17)



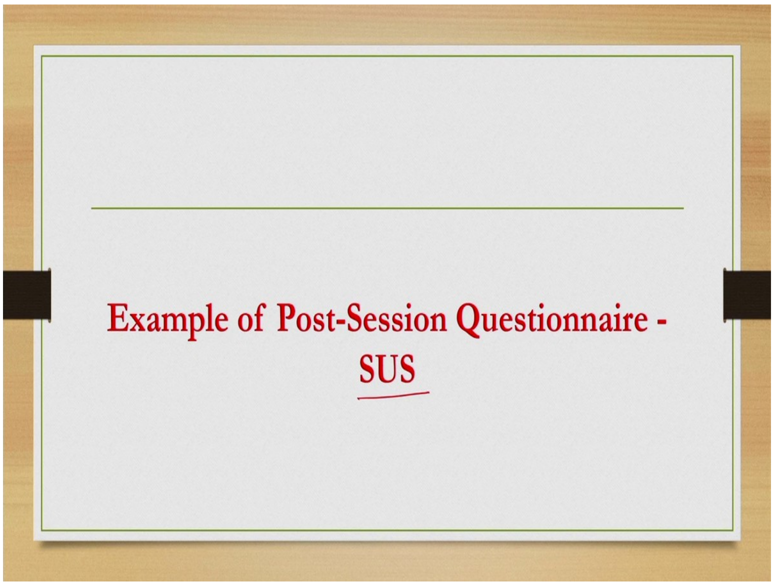
Basic Idea

- Use of the feedback
  - By considering average rating for each statement
  - Or average rating for all statements taken together

And, how we can use the feedback receipt through ratings? We can consider the average ratings for each statement that is one way. The other way is we can consider average rating for all the statements taken together to come to a conclusion about the overall usability of that particular system with respect to the particular task.

So, the idea is that if we get these ratings for each task at the end of each task, then, at the end of the session we can actually go through these ratings, find out the task for which the rating is very poor. And, for those tasks we can actually try to figure out why the ratings are poor, why the users felt that the system is not usable, which components of the interface were part of the task and how to improve those components. So, in other words with these ratings, the designers will be able to identify specific problem areas in the overall design.

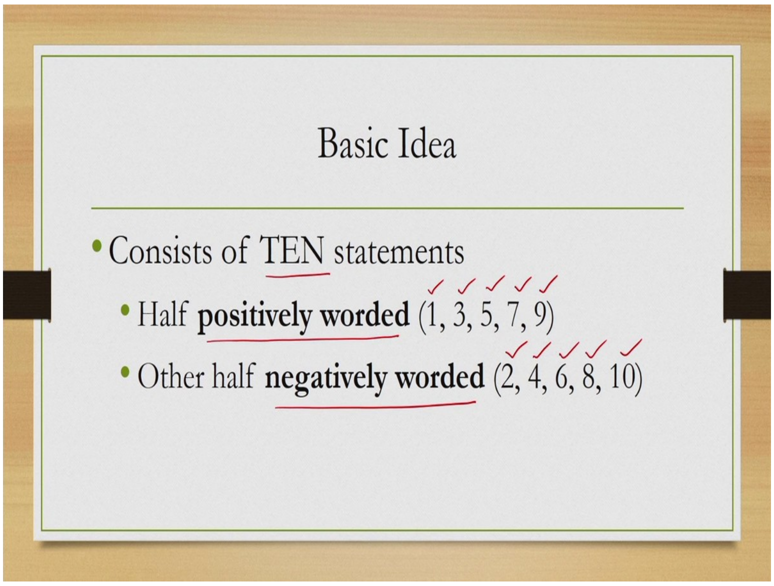
(Refer Slide Time: 18:44)



Example of Post-Session Questionnaire - SUS

Now, that is about post task questionnaire example. Let us now move to post session questionnaire example that is SUS or System Usability Score.

(Refer Slide Time: 19:00)



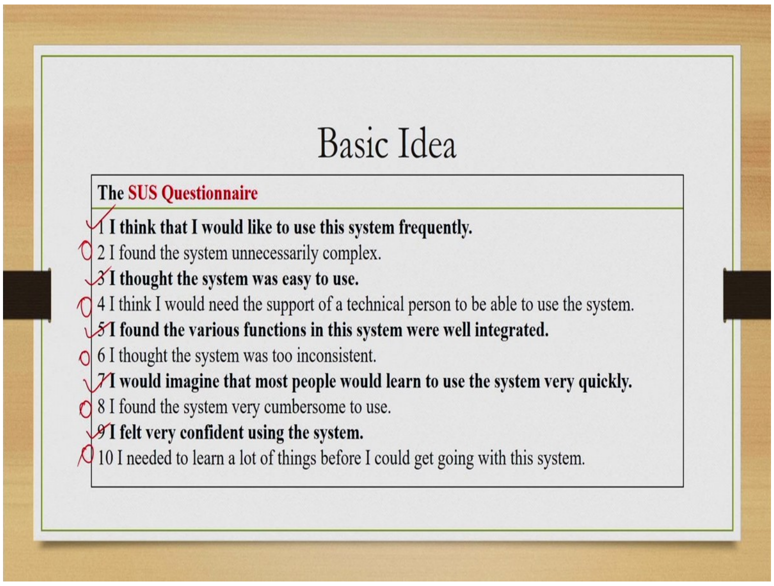
Basic Idea

- Consists of TEN statements
  - Half **positively worded** (1, 3, 5, 7, 9)
  - Other half **negatively worded** (2, 4, 6, 8, 10)

Now, SUS consists of ten statements. Earlier in ASQ, we had three statements; here in SUS, we have ten statements. One important thing is that among the ten statements half or 5 bar positively worded. These are statement numbers 1, 3, 5, 7 and 9 will soon see. And, other half or the remaining 5 statements are negatively worded these are statement numbers 2, 4, 6, 8 and 10.

Let us now see these ten statements to understand these concepts.

(Refer Slide Time: 19:50)



Basic Idea

The **SUS Questionnaire**

✓ **1 I think that I would like to use this system frequently.**
O 2 I found the system unnecessarily complex.
✓ **3 I thought the system was easy to use.**
O 4 I think I would need the support of a technical person to be able to use the system.
✓ **5 I found the various functions in this system were well integrated.**
O 6 I thought the system was too inconsistent.
✓ **7 I would imagine that most people would learn to use the system very quickly.**
O 8 I found the system very cumbersome to use.
✓ **9 I felt very confident using the system.**
O 10 I needed to learn a lot of things before I could get going with this system.

These are the ten statements for SUS questionnaire. The first statement is I think that I would like to use this system frequently. As you can see this is a positively worded statement.

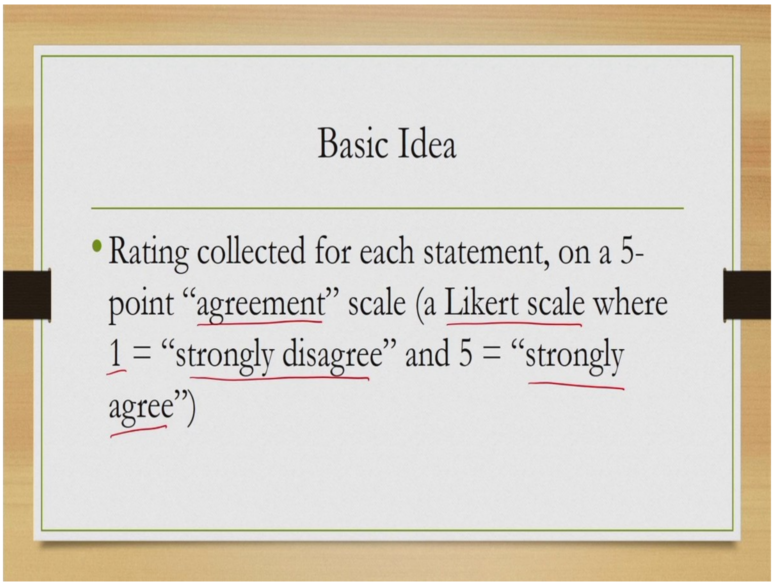
Second statement is I found the system unnecessarily complex, this is a negatively worded statement. Third one is I thought the system was easy to use again, this is a positively worded statement. And, forth is I think I would need the support of a technical person to be able to use the system which is a negatively worded statement.

Similarly, the fifth statement is a positively worded statement which says I found the various functions in this system where well integrated whereas, the sixth statement is negatively worded and says, I thought the system was too inconsistent. Seventh statement is I would imagine that most people would learn to use the system very quickly. Whereas eighth is a negatively statement, which says I found the system very cumbersome to use. Ninth statement is I felt very confident using the system and tenth is I needed to learn a lot of things before I could get going with this system.

So, in these ten statements, the positively worded statements are shown in bold font and the negatively worded statements are shown in regular font.

So, 1, 3, 5, 7 and 9 are positively worded and 2, 4, 6, 8, 10 are negatively worded together they constitute the SUS questionnaire.

(Refer Slide Time: 21:55)

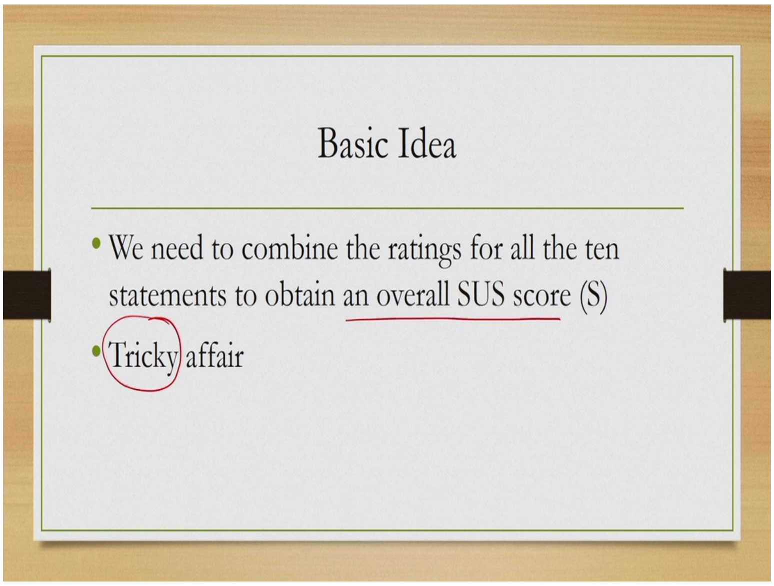


Basic Idea

- Rating collected for each statement, on a 5-point "agreement" scale (a Likert scale where 1 = "strongly disagree" and 5 = "strongly agree")

So, for each question or for each statement in the questionnaire, the feedback is collected on a rating scale which is a 5-point rating scale, also called an agreement scale which is a Likert type scale, where 1 indicates strongly disagree and 5 indicates strongly agree. So, at the end of a session each participant in the study is given this questionnaire and their ratings are collected for each of these ten statements.

(Refer Slide Time: 22:43)

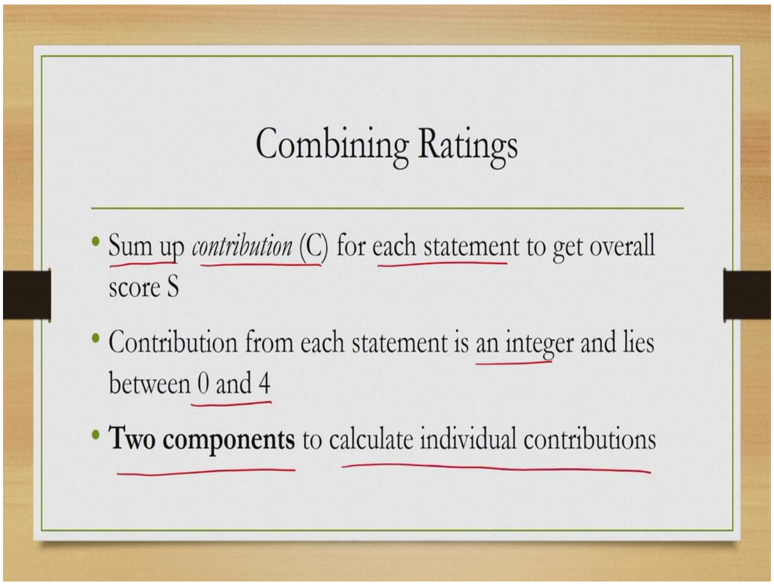


Basic Idea

- We need to combine the ratings for all the ten statements to obtain an overall SUS score (S)
- Tricky affair

Then, we need to combine these ratings. And, after combination we get a score which is an overall SUS score which indicates the usability of the system. Now, this the combining the ratings is not a straight forward thing, it is tricky and will go into the details of this combination process. This is unlike the ASQ rating where we can simply take some average and that is all, but here we required to do some other things.

(Refer Slide Time: 23:20)

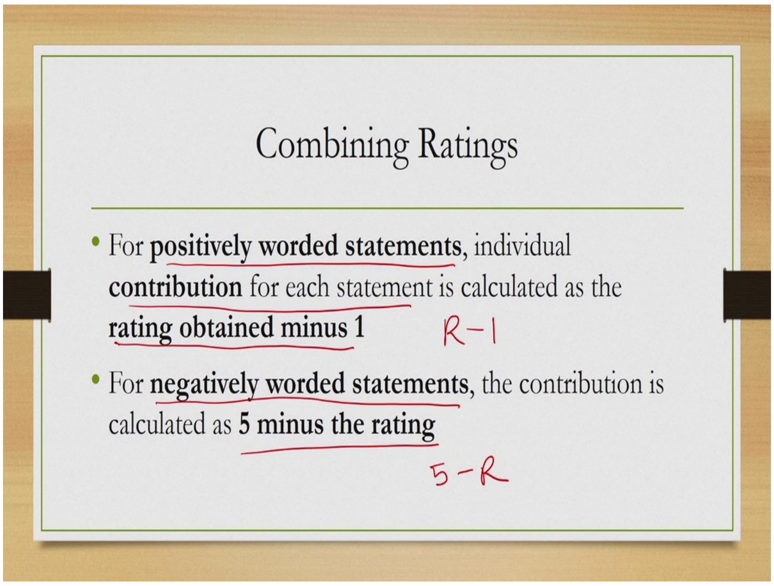


Combining Ratings

- Sum up *contribution* (C) for each statement to get overall score S
- Contribution from each statement is an integer and lies between 0 and 4
- **Two components** to calculate individual contributions

First thing that we need to do is to sum up the contributions for each statement. In order to do that we should be able to find out the contributions for each statement which is an integer and lies between 0 to 4. And, these contributions are actually of two types. So, accordingly, there are two components to calculate these individual contributions.

(Refer Slide Time: 24:04)



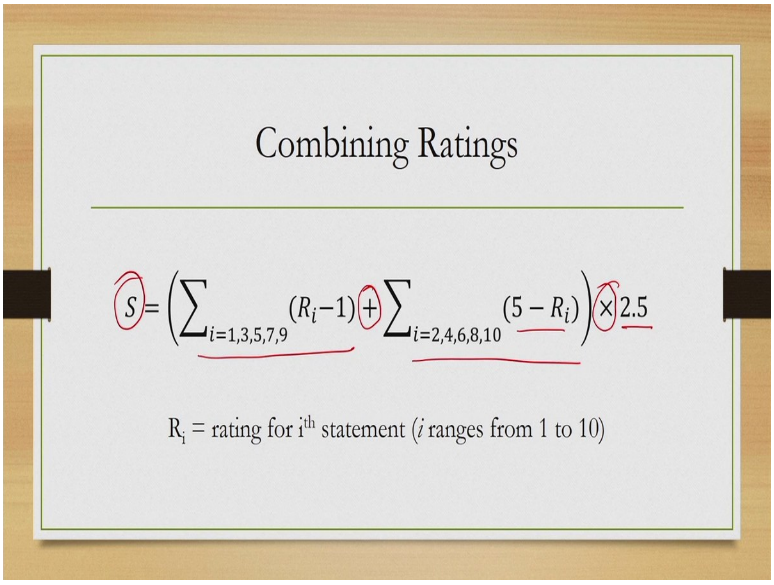
Combining Ratings

- For **positively worded statements**, individual **contribution** for each statement is calculated as the **rating obtained minus 1**   R-1
- For **negatively worded statements**, the contribution is calculated as **5 minus the rating**

5-R

For positively worded statements, individual contribution for each statement can be calculated as the rating obtained minus 1. So, if R indicates the rating then for each statement, we will get the individual contribution by the formula R minus 1. For negatively worded statements, the contribution is calculated as 5 minus the rating. So, it is 5 minus R.

(Refer Slide Time: 24:44)



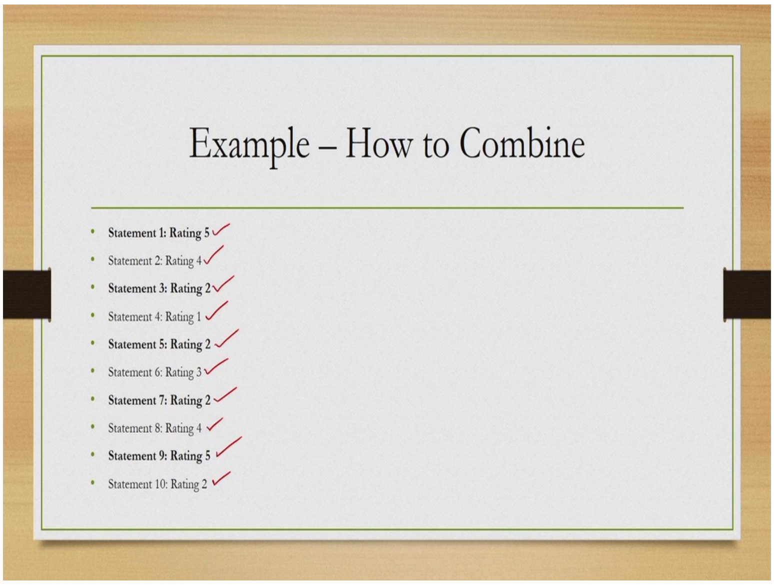
Combining Ratings

$$S = \left(\underline{\sum_{i=1,3,5,7,9}(R_i - 1)} + \underline{\sum_{i=2,4,6,8,10}(5 - R_i)}\right) \times \underline{2.5}$$

R_i = rating for ith statement (*i* ranges from 1 to 10)

The equation shows the overall calculation. So, here S is the SUS score. There are two components for positively worded and for negatively worded. R i indicates the rating for i'th statement where i ranges from 1 to 10. Now, for positively worded there are 5 individual components all these are summed together. Similarly, for negatively worded statements there are 5 components and for each we calculate the contribution with this equation and then we sum them together. Then, these two components are added. So, we calculate an overall sum and then we multiply this sum with the factor 2.5 to get the score.

(Refer Slide Time: 25:48)

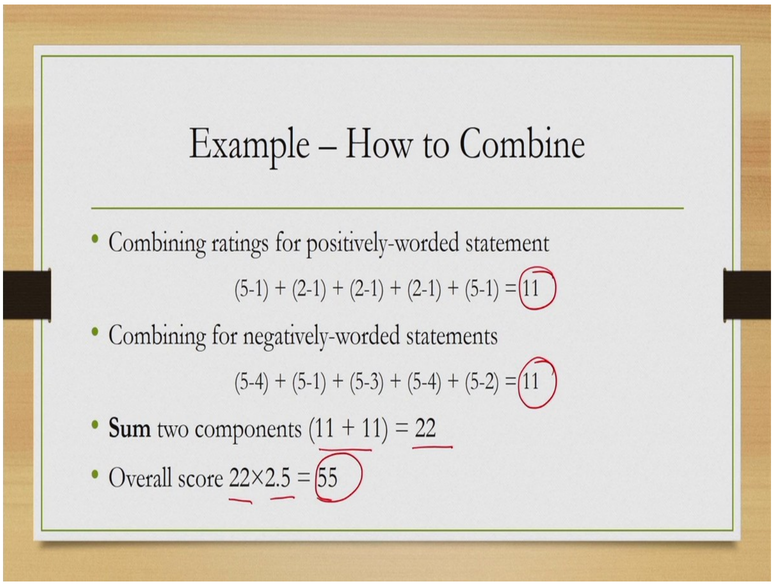


Example – How to Combine

- **Statement 1: Rating 5** ✓
- Statement 2: Rating 4 ✓
- **Statement 3: Rating 2** ✓
- Statement 4: Rating 1 ✓
- **Statement 5: Rating 2** ✓
- Statement 6: Rating 3 ✓
- **Statement 7: Rating 2** ✓
- Statement 8: Rating 4 ✓
- **Statement 9: Rating 5** ✓
- Statement 10: Rating 2 ✓

As, an example let us consider these ratings that we have received for a particular system. For statement 1, we got a rating of 5, for 2 it is 4, for 3 it is 2, for 4 it is 1, for 5 it is 2, for 6 it is 3, for 7 it is 2, for 8 it is 4, for 9 it is 5 and for statement 10, we receive the rating of 2. Now, the ratings for positively worded statements are shown in bold and the ratings for negatively worded statements are shown in regular font. So, with these ratings, we can use the equation to calculate the overall SUS score.

(Refer Slide Time: 26:40)



Example – How to Combine

- Combining ratings for positively-worded statement

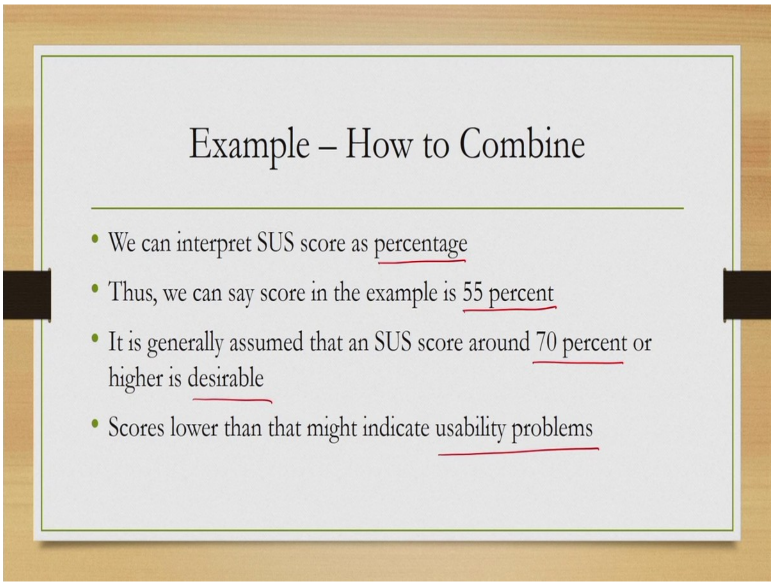
  $(5\text{-}1) + (2\text{-}1) + (2\text{-}1) + (2\text{-}1) + (5\text{-}1) = 11$
- Combining for negatively-worded statements

  $(5\text{-}4) + (5\text{-}1) + (5\text{-}3) + (5\text{-}4) + (5\text{-}2) = 11$
- **Sum** two components $(11 + 11) = 22$
- Overall score $22 \times 2.5 = 55$

So, first we find out the individual components for the positively worded statement using the equation R minus 1; so, rating minus 1 and we get 11. Then, we find out the overall contribution by the negatively worded statements by summing up the individual contribution for each statement, using the equation 5 minus R.

So, here we get the same value incidentally for this component as well that is 11, then we add the two components and get 22 within multiply it with this factor 2.5 to get this overall score 55.

(Refer Slide Time: 27:39)



Example – How to Combine

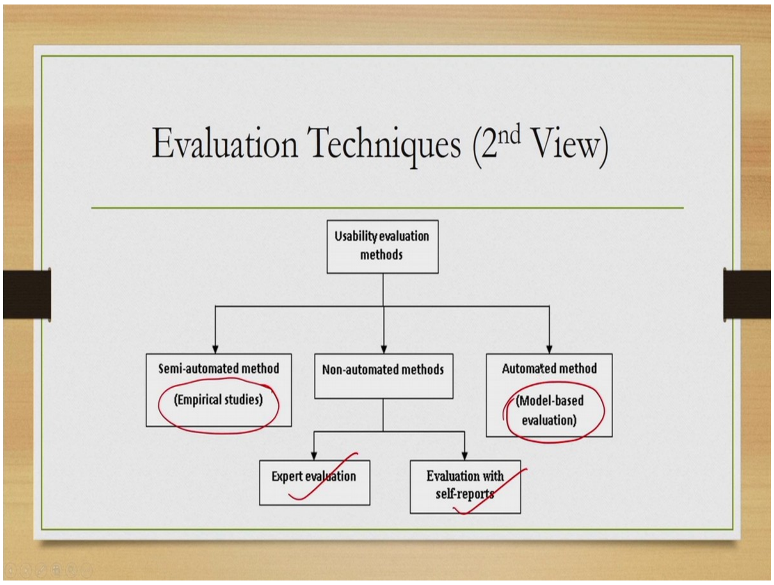
- We can interpret SUS score as percentage
- Thus, we can say score in the example is 55 percent
- It is generally assumed that an SUS score around 70 percent or higher is desirable
- Scores lower than that might indicate usability problems

Now, what 55 tells us? We can interpret the score as percentage. So, when we say that we got a score of 55, we can say that, the score is 55 percent. And, it is generally assumed it has been found that a score around 70 percent or more is desirable.

So, if we get a score which is less than this value, then it indicates there may be some usability problems. Where exactly the problem is? It cannot tell which is any of the characteristics of a post session feedback mechanism, but it can tell that there are some problems and the designer has to go through the design again minutely to find out the source of the problems.

Now, that is about evaluation with self-report. So, we have discussed two evaluation methods; one is expert evaluation, one is evaluation with self-report where we involve the users.

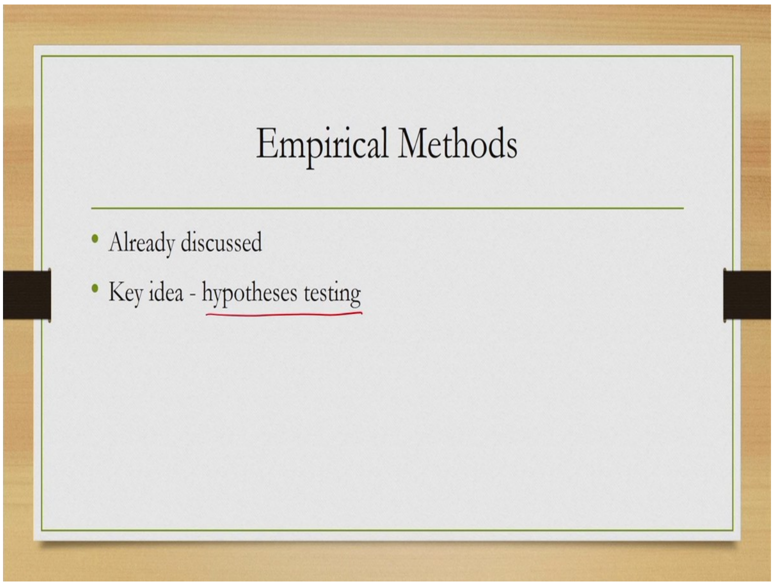
(Refer Slide Time: 28:48)



Evaluation Techniques (2nd View)

So, in the hierarchy of evaluation methods so, we have discussed expert evaluation and evaluation with self-report. In expert evaluation, we discussed about cognitive walkthrough and heuristic evaluation, which self-reports we talked about post task and post session questionnaire and feedback based on rating scales and we discussed couple of standard questionnaires, namely the ASQ and the SUS. What is left is the other methods, namely the empirical studies and model based evaluation. So, we have already discussed this methods before. Let us just quickly recap what we have discussed.

(Refer Slide Time: 29:34)

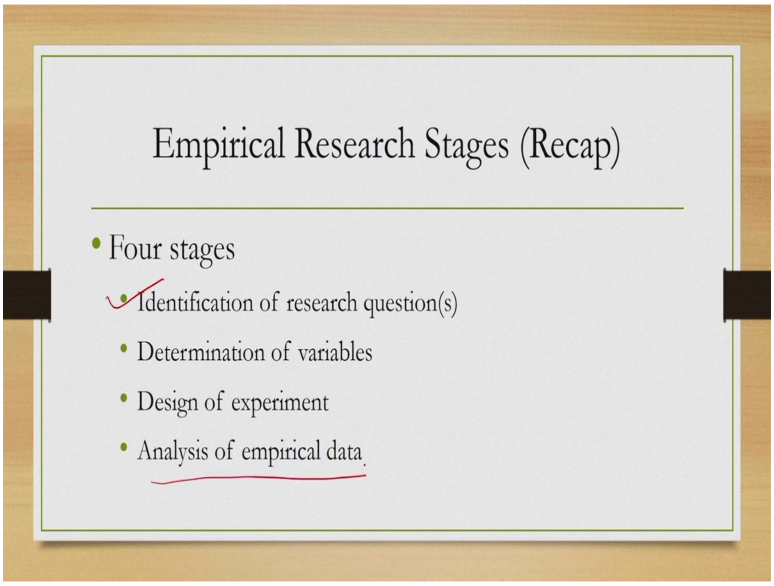


Empirical Methods

- Already discussed
- Key idea - hypotheses testing

Now, when we talk of empirical evaluation method, we have essentially referred to hypothesis testing.

(Refer Slide Time: 29:46)



Empirical Research Stages (Recap)

- Four stages
  - Identification of research question(s)
  - Determination of variables
  - Design of experiment
  - Analysis of empirical data

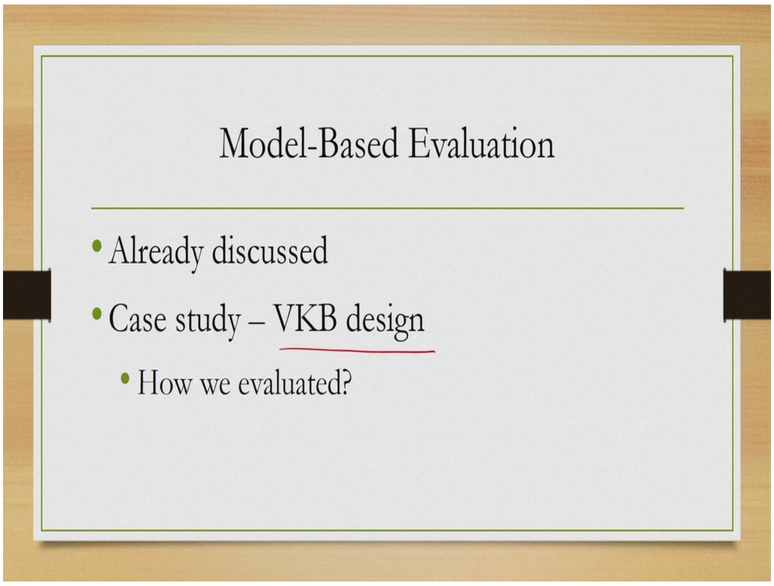
As we discussed in details in a previous lecture, it involves four stages. So, we start with identification of research questions and this is followed by framing of hypothesis from each question, then we determine the variables for which we want to observe and record data.

So, we determine dependent variables, independent variables, control variables and confounding variables, then we go for the design of our experiment. So, in the design several things are taken into account namely how many participants, what kind of tasks, in which order the tasks are to be given, whether the experiment should be within subject or between subject and so on.

And, finally, we go for analysis of empirical data essentially this stage refers to the analysis of data for testing of statistical significance. So, that we can refute the null hypothesis and support alternative hypothesis. And, these alternative hypothesis is essentially refers to a quality which supports usability.

So, the overall objective of empirical study is to evaluate the usability of the system in terms of hypothesis testing, or in terms of supporting alternative hypothesis through statistical means by collecting data in a controlled environment.

(Refer Slide Time: 31:27)



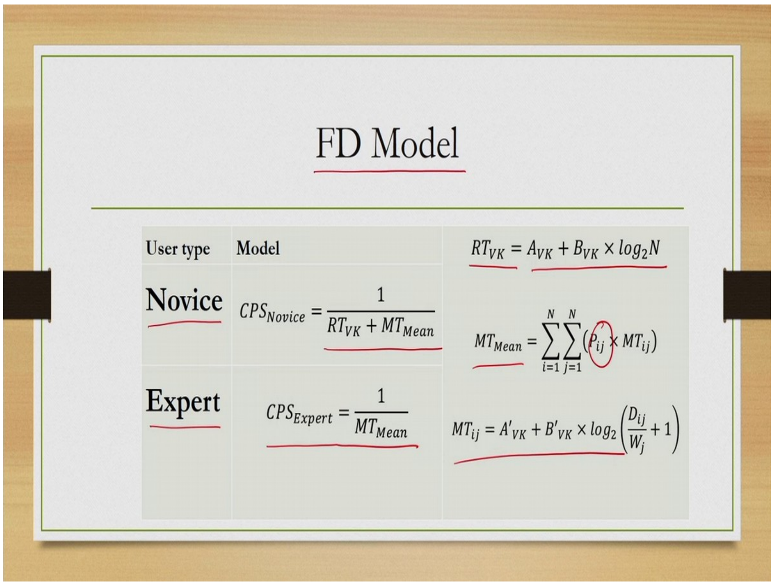
Model-Based Evaluation

- Already discussed
- Case study – VKB design
  - How we evaluated?

The other evaluation technique is model based. Now, we already have discussed this in an earlier lecture like empirical studies and there we discussed about case studies on virtual keyboard design.

Let us just quickly recap, how we can evaluate a design using models.

(Refer Slide Time: 31:56)



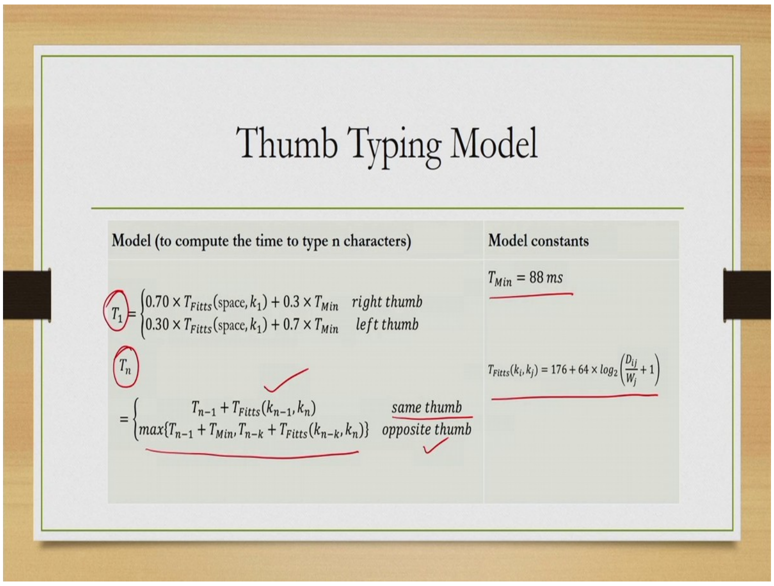
FD Model

User type	Model	$RT_{VK} = A_{VK} + B_{VK} \times log_2 N$
Novice	$CPS_{Novice} = \frac{1}{RT_{VK} + MT_{Mean}}$	$MT_{Mean} = \sum_{i=1}^{N}\sum_{j=1}^{N}(P_{ij} \times MT_{ij})$
Expert	$CPS_{Expert} = \frac{1}{MT_{Mean}}$	$MT_{ij} = A'_{VK} + B'_{VK} \times log_2\left(\frac{D_{ij}}{W_j} + 1\right)$

So, we talked of two models; one is the Fitts' diagram model this was developed to evaluate performance of single finger typing on a mobile device using virtual keyboards. There the model is separate for a novice user and then expert user for novice, user we can compute performance in terms of character per second using this equation where RT is choice reaction time and MT is the average movement time between keys on the layout. Whereas, for expert we can compute performance by this expression, where CPS is the character per second MT mean is the average movement time between keys on a layout.

And, the choice reaction time or the reaction time can be computed using the hick Hyman law, the average movement time can be computed with help of the Fitts' law and the diagram probability distribution found from a corpus. So, what we evaluate? We evaluate the performance. And, in this case the performance is represented in terms of characters per second entered by the typist, or character per second which is likely to be entered by an average typist which as we have discussed is related to the measure of usability.

(Refer Slide Time: 33:33)



Thumb Typing Model

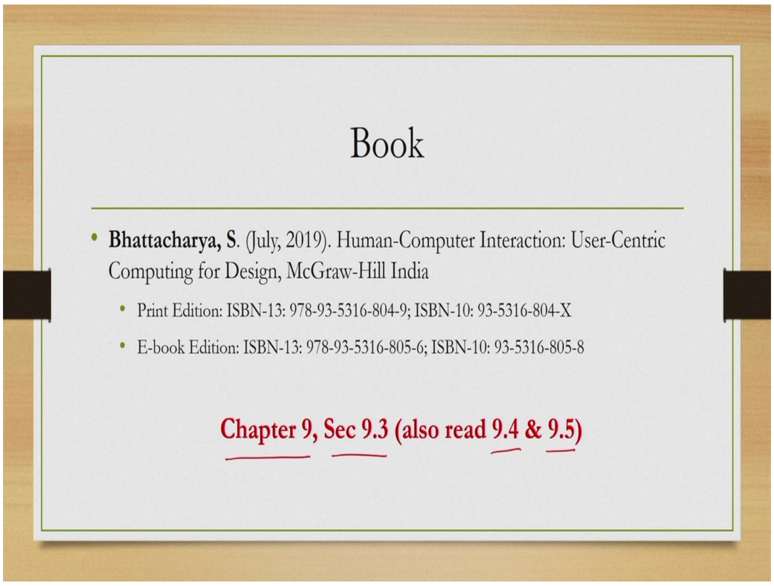
Model (to compute the time to type n characters)	Model constants
$T_1 = \begin{cases} 0.70 \times T_{Fitts}(\text{space}, k_1) + 0.3 \times T_{Min} & right\ thumb \\ 0.30 \times T_{Fitts}(\text{space}, k_1) + 0.7 \times T_{Min} & left\ thumb \end{cases}$ $T_n = \begin{cases} T_{n-1} + T_{Fitts}(k_{n-1}, k_n) & same\ thumb \\ max\{T_{n-1} + T_{Min}, T_{n-k} + T_{Fitts}(k_{n-k}, k_n)\} & opposite\ thumb \end{cases}$	$T_{Min} = 88\,ms$ $T_{Fitts}(k_i, k_j) = 176 + 64 \times log_2\left(\frac{D_{ij}}{W_j} + 1\right)$

Similarly, we have talked about thumb typing model for mobile typing. This model is somewhat different. So, it is a combination of equations and algorithms to compute the overall performance. The equations are recursive T n indicates the time required to type using 2 thumbs, n characters which is represented using these 2 equations it has 2 components. If you are using the same thumb the top component is applicable.

If you are using opposite thumb then the bottom component is applicable. T 1 is the time to type the first character and there are some model parameters which are found empirically. And, once we are able to compute the time to enter n characters using that we can compute overall performance in terms of characters per second through an algorithm.

So, these two case studies for model based evaluation indicates that, we can apply the models to find out the usability, but in a limited sense only with respect to some components of usability, but eventually we have to go for thorough empirical study at the end.

(Refer Slide Time: 35:26)



Book

- **Bhattacharya, S.** (July, 2019). Human-Computer Interaction: User-Centric Computing for Design, McGraw-Hill India
  - Print Edition: ISBN-13: 978-93-5316-804-9; ISBN-10: 93-5316-804-X
  - E-book Edition: ISBN-13: 978-93-5316-805-6; ISBN-10: 93-5316-805-8

Chapter 9, Sec 9.3 (also read 9.4 & 9.5)

Whatever we have discussed today can be found in this book. So, in today's lecture as well as the previous lecture we have covered different evaluation techniques. All these techniques can be found in this book, you are advised to refer to chapter 9. For today's lecture on evaluation with self-report, you can refer to section 9.3. However, you are also advice to go through section 9.4 and 9.5 for a better understanding of model based design that is all for today.

Thank you and goodbye.