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

Lecture – 37
Empirical Usability Evaluation - 1

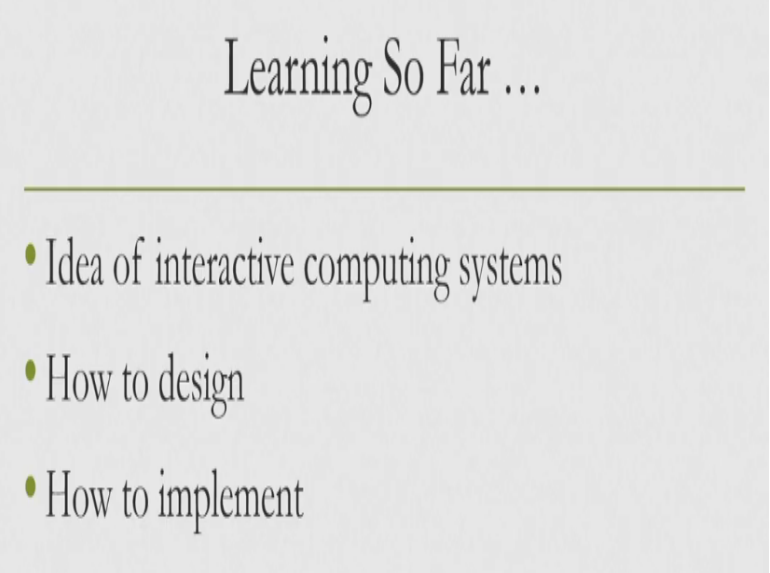
Hello and welcome to the NPTEL MOOCS course on design and implementation of human-computer interfaces. We are in lecture number 31 where we are going to start a new topic. As the convention we are following in the previous lectures, before we start we are going to talk about briefly what we have learned so far and where we are now in the context of this overall course. If you may recollect, we are discussing about interactive system development.

So, when we use the term human-computer interfaces essentially what we are referring to is development of interactive systems. Interactive systems are those systems that are supposed to be used by the lay persons those who are not experts in technology. Also, it may be recalled at this stage that when we talk of system, our primary intention is to refer to software systems rather than any hardware.

So, when we are talking about interactive systems or human-computer interfaces as the title of this course suggests, we are essentially referring to the development of interactive software that are to be used by lay persons who are not technology experts, who are not experts in the knowledge of how the technology works in the background and who are also not interested in knowing or having that knowledge.

What they are interested in is primarily to get their work done. For those type of users, we require special type of considerations to build the software.

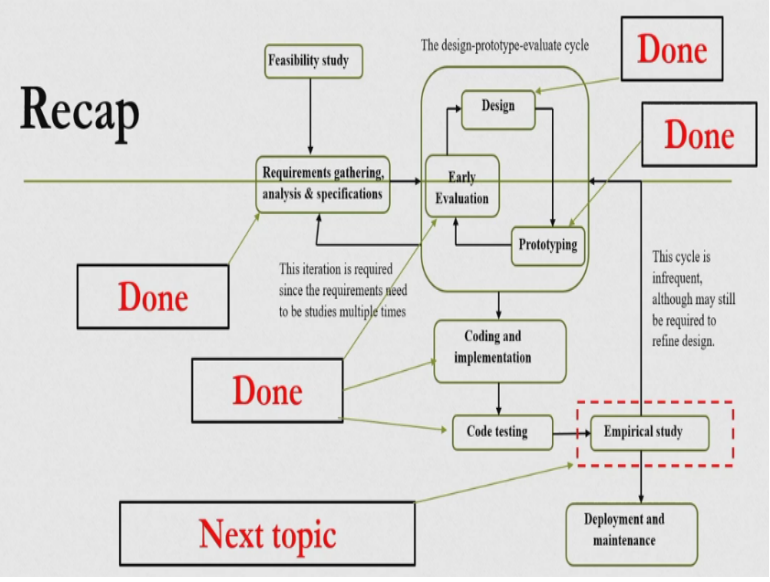
(Refer Slide Time: 02:31)



Now, in the previous parts of this course, in the previous 29 lectures and the few case studies that we have covered so far, we learned in details about the idea of interactive systems, what are those and how those are different from other systems. We have also learned how to identify requirements for those systems and how to translate those requirements into the design of a system and then how to implement that as a software.

Now, we are going to learn about an important aspect of the development of such type of systems that is how we can ensure that the system that we have developed meets the expectations, the needs, the goals of the end users. In order to ensure that the system actually meets the needs and expectations of the end users, we take help of a systematic system development process. Now in this process, there are several stages.

(Refer Slide Time: 03:56)



So, we started our discussion with the requirement gathering stage that is the first stage. At the end of the stage, we identify the requirements that are there, that are to be considered for making the system acceptable to the end users. At the end of this stage, we produce a document called SRS or software requirements specification document. After that, we have learned about the design prototype evaluate cycle. In this cycle, there are three stages.

Design stage, prototype building stage and evaluation stage. So, in the design stage we primarily focus on design of the interface and interaction which is the primary component from the point of view of the user. So to a user an interactive software essentially refers to the interface and interaction of that software rather than the internal design of the software. This is followed by prototyping.

So once the design is done, we prototype it and then we evaluate it with experts. This process continues till we arrive at a stable interface design where not many issues are identified or not many issues remain. At the end of each stage, we produce a document. There can be a design document, prototype document as well as an evaluation report of the prototypes. So, once we finalize interface design, we go for system design and produce a system design document.

If you may recall it system design can be done either in a functional oriented approach or procedural approach where we use languages such as DFDs and ERDs to create the document. In case of object-oriented approach, we can use languages such as UML to produce the design document. Once the design is done, then we implement the system using coding and there we have seen what kind of coding practices should be followed for a good implementation.

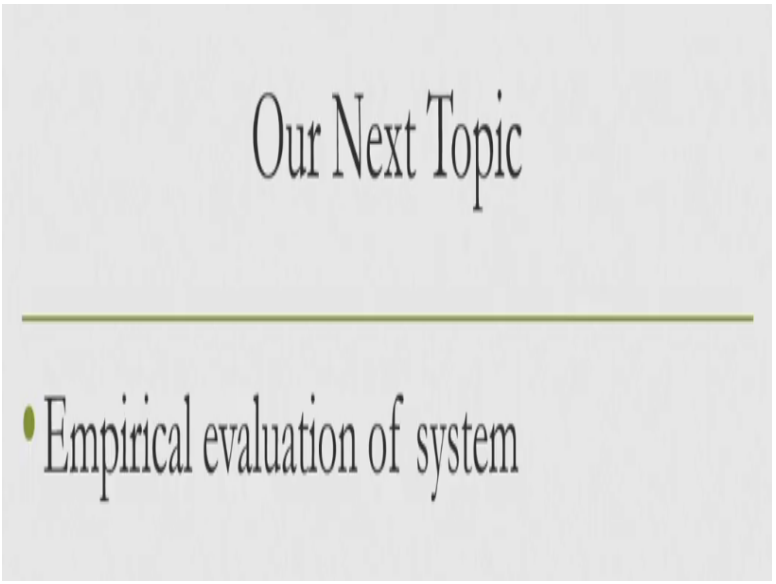
It will be recalled that after design activities for the system gets over, we need to produce the design document. Then this document is used for implementing the system. At the end of this implementation stage, we produce the code which may be documented with comments and also some manuals to understand that code. After that, we have covered in details the testing of the code, the purpose of testing is to ensure that there are no bugs or as little bugs as possible.

And we have learned in details different techniques which include review-based testing. We have learned in details different testing methods which include review-based testing, black

box testing and white box testing. So, each of these methods can result in a testing report as we have seen in our earlier lectures. Now, let us come to the subject matter of this lecture in the context of this interactive system development lifecycle.

So, we are going to now start discussion on the next stage of this lifecycle that is empirical study. This is going to be our next topic and we will devote this lecture as well as subsequent few lectures on explaining this stage.

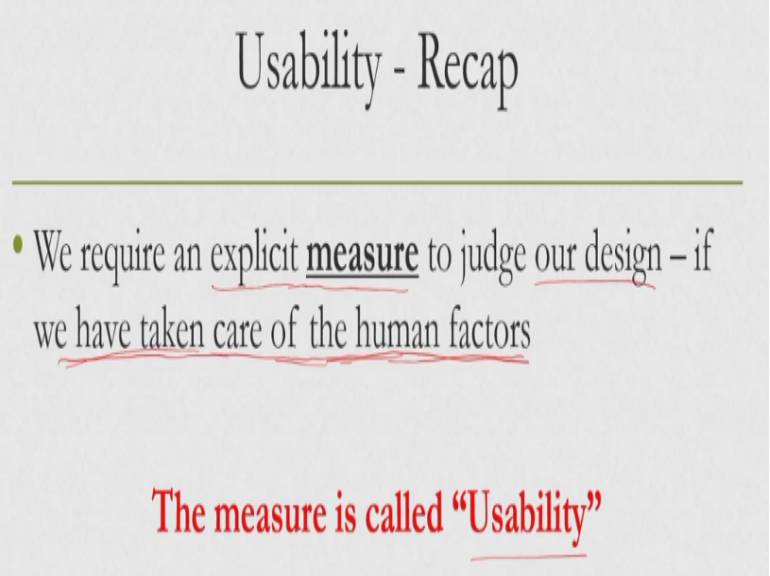
(Refer Slide Time: 07:47)



So, empirical evaluation of the system is going to be our next topic. Just to put it into perspective, so earlier we have seen evaluation of the system from the point of view of the code. So, we have written a code and we tested it using different methods so that ensured that the code is bug free to the extent possible. The purpose of empirical evaluation on the other hand is not to identify bugs in the code.

Rather it is related to the objective of ensuring that the product is acceptable to the end users, the product is designed in a way such that it meets the expectations and needs and goals of the end users. In other words, the objective is to ensure usability of the product. So, that is the major purpose.

(Refer Slide Time: 08:44)



Now, let us quickly recap the idea of usability. When we say that we perform empirical evaluation of the product that we are developing or rather the interactive software that we are developing what we really mean, what is usability? If you may recollect, we discussed usability at the beginning of this course in the first week itself, the idea of usability was discussed in details.

Since usability is going to be the main objective for this stage, so let us quickly recap the idea again here. What is usability? Briefly we can refer to it as an explicit measure, measure for what? So, it is an explicit measure to judge our design if we have taken care of the human factors. So, this is the important consideration when we are talking of interactive systems. As I just mentioned, it is meant to be used by lay persons who are not technology experts.

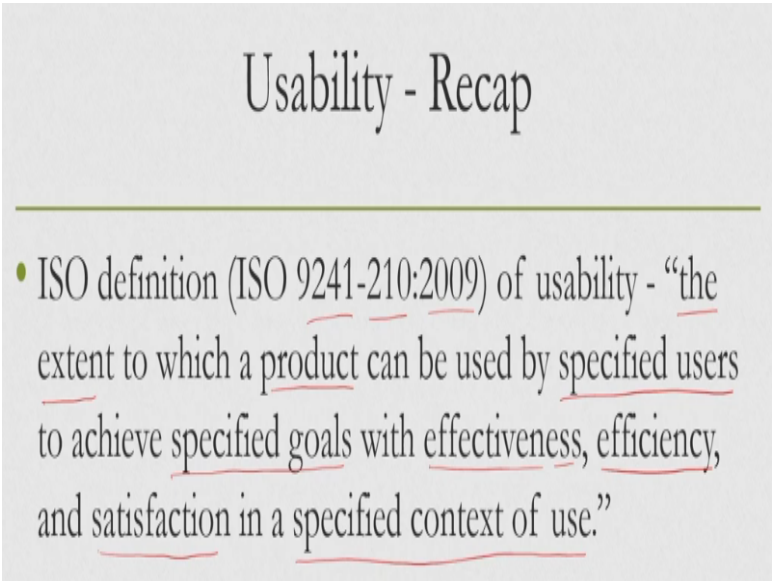
They are not interested actually to learn about the technology behind the working of the system. What they are interested in is basically to get their work done. For example, if a traveller wants to book a train ticket and he uses a ticket booking app, the traveller is not interested to know how the app works, rather the traveller is interested to achieve the goal of booking a ticket using the interface provided by the system as well as the interaction supported by the ticket booking app.

So, whether the traveller will be able to do that or not depends on whether we have considered the human factors in our design of the ticket booking app. Now, those human factors have to be taken care of in the design of any interactive software. Now, the idea of

usability actually is to give us a measure to know whether those considerations are taken into account, so that is the idea of usability.

And an explicit measure to judge our design from the point of view of a consideration of human factors that are relevant for our system. It may be noted that a product which we found to be usable for a specific group of system need not be usable for another group because the human factors for the other group may be different than the group we are considering, so that is also very important.

(Refer Slide Time: 11:27)

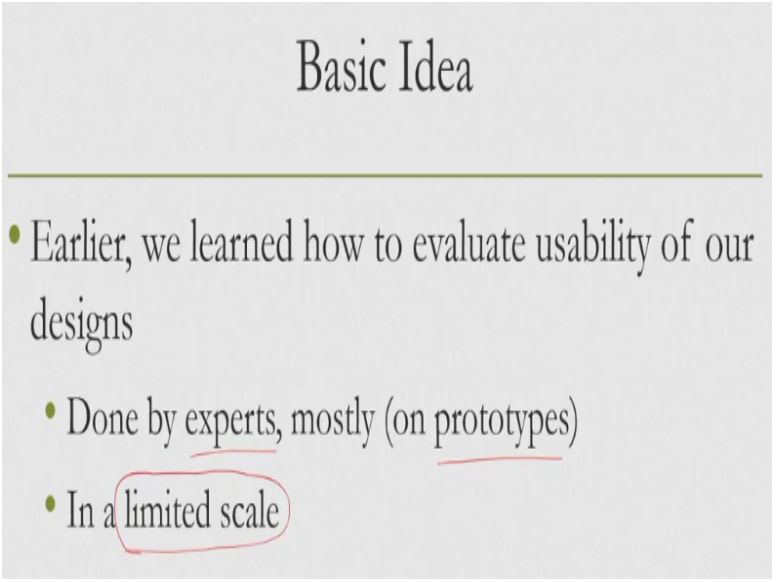


Let us quickly recall the definition of usability, definition provided by the International Standards Organization, the standard number 9241-210:2009. So, according to this definition of usability, usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. We have already covered this definition in details including the significance of the terms mentioned here.

Just to recollect, there are three things mentioned that is usability is a measure to know whether a product can be used by a specific group of users to achieve specific set of goals in a specific usage context with effectiveness, efficiency and satisfaction. So, these three terms effectiveness, efficiency and satisfaction together constitute the measure but equally important to notice that we measure usability against a specific what context, specific group of users and specific set of goals.

So, as I said whenever we are talking of usability, we should keep these things in mind that usability is defined in the context of specific work contexts, specific group of users and specific set of goals. So, same group of users may find a product unusable if the set of goals change or the work context change. So, these things are key considerations while we are talking about usability.

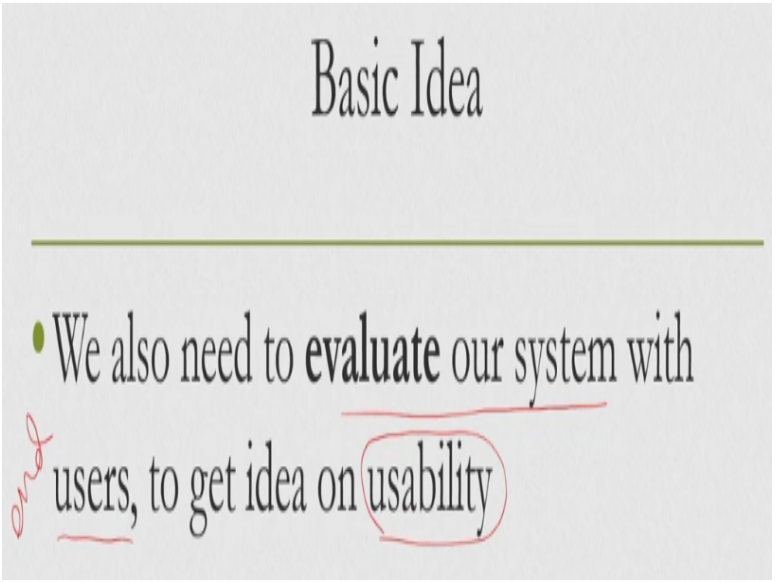
(Refer Slide Time: 13:35)



Now with that background knowledge which we just recollected, let us come to the main part of this topic, empirical study. So, our objective is then to learn whether a product is usable, earlier we have briefly talked about measuring usability in the context of prototype evaluation. So we said that once we build a prototype, we employ a set of quote unquote experts, who are essentially domain experts, not the technology experts to measure usability of the product in a quick manner.

But that evaluation methods were as pointed out done by experts mostly on prototypes, in fact only on prototypes and in a limited scale. So, it is a limited kind of measurement ad it may not give us true measure of usability when the product is to be used by end users. Experts are limited in number and they may not reflect fully the behaviour of the end users. So, what we need is we have to go for full scale evaluation.

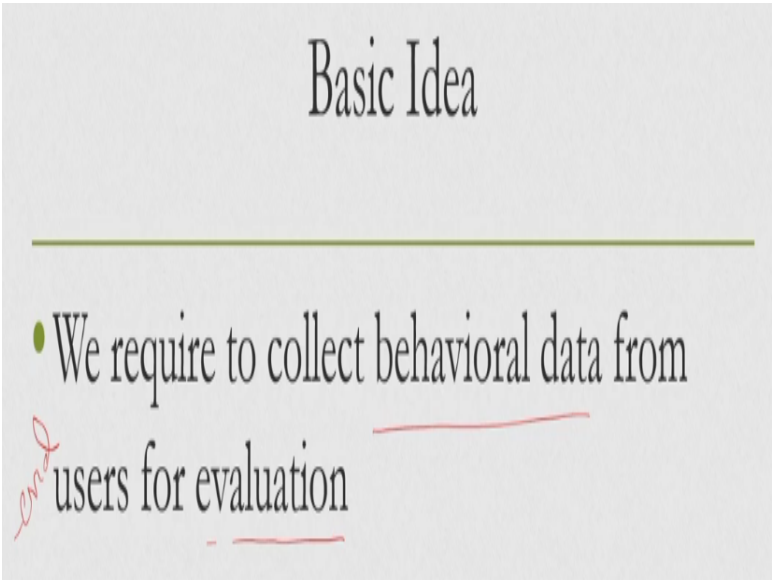
(Refer Slide Time: 15:01)



We need to evaluate our systems with end users, so here users mean end users. Now, the term end user is very important here. In the context of software development, we have to keep in mind that sometimes the task of developing the software can be given to a developer by a customer who is not necessarily the end user. So, customer is the user of the software, but there can be another group called end user, so here our focus will be only on end users.

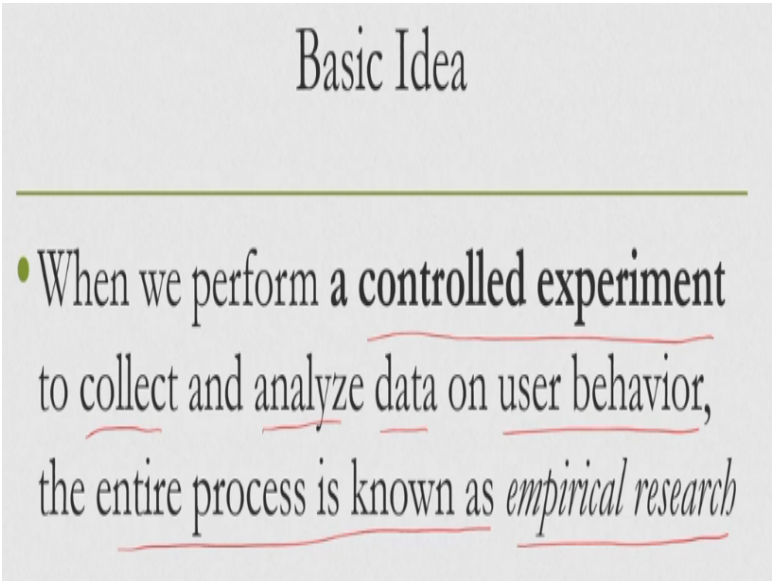
So, what we need is, we need to evaluate our system with end users to get an idea of usability of the final software. This is in contrast to the earlier evaluation where we evaluated our prototypes with experts on a limited scale, now here we are talking of evaluating the whole system with end users, actual users in a more rigorous manner.

(Refer Slide Time: 16:09)



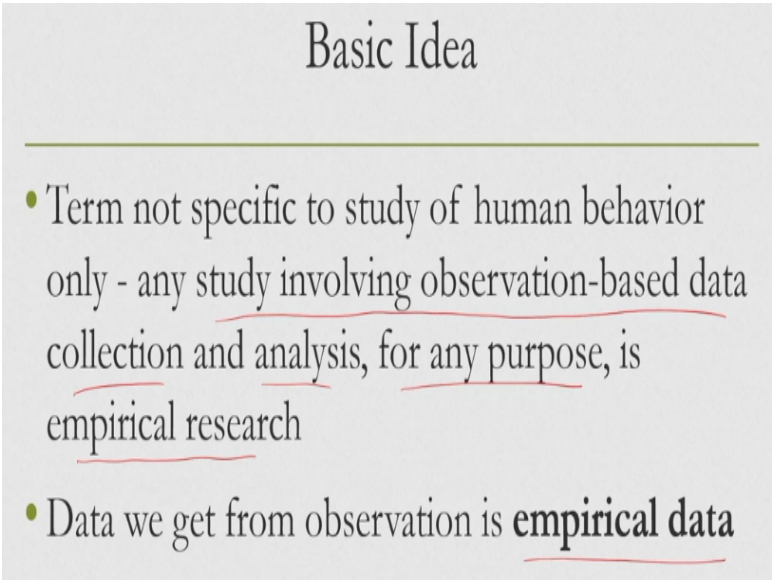
In order to do that, what we need to do? We require to collect behavioural data from the end users. Again, I would like to emphasize on the point end users, we need to collect behavioural data from the end users for the purpose of usability evaluation that is what we require.

(Refer Slide Time: 16:31)



So, essentially what we need to do is we need to collect data from end users, the data is behavioural data. When we perform a controlled experiment to collect and analyse data on user behaviour, the entire process is known as empirical research. So, when we are using the term empirical study or empirical research of end users for usability, essentially we are referring to a general process, a controlled experiment we are carrying out to collect behavioural data from end users. So, this is in general known as empirical research.

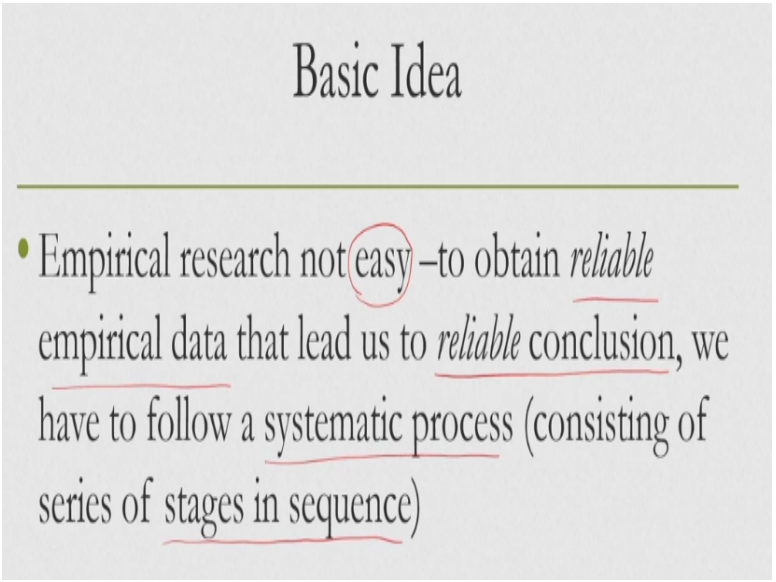
(Refer Slide Time: 17:20)



It may be noted here that this term empirical research is not specific to the evaluation of interactive systems only. So, it is not specific to this human-computer interfaces, it is a generic concept, any study involving observation-based data collection and analysis for any purpose is empirical research. So, when we are actually going for collecting data based on observation and then analysing that data that is called empirical research.

So, it is not only collecting end user data for usability evaluation of human-computer interfaces, it is a generic concept which we are applying for our purpose of usability evaluation of human-computer interfaces. Now, the data that we get from such observations is called generally empirical data. So, that is a common term a generic term rather than term specific to our context here that is evaluation of usability of human computer interfaces.

(Refer Slide Time: 18:44)



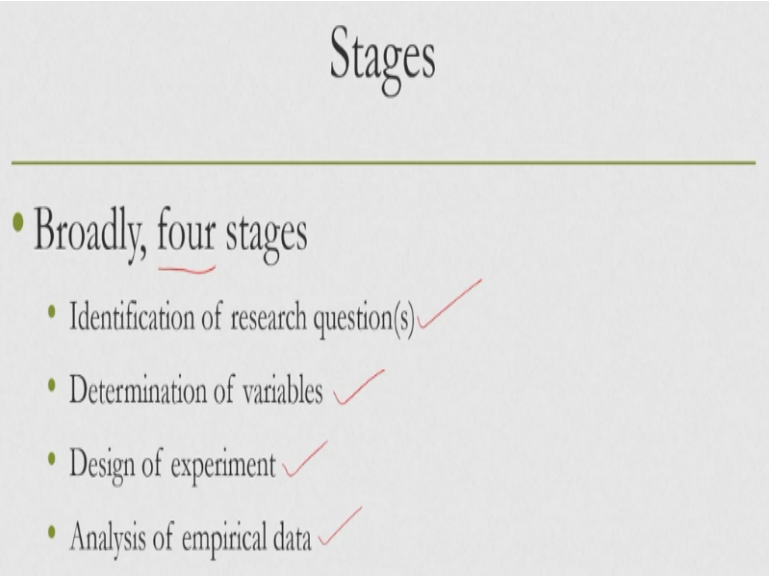
One problem is or rather one issue with empirical research is it is not easy. You may find or you may think what is the big deal here? We simply ask some users about our system that we have developed, we may take their feedback, we may ask them whether they find it usable or not directly and based on their response, we may decide whether it is usable or not, but it is not as simple as that.

To obtain reliable empirical data that leads us to reliable conclusion, we have to follow a very systematic process which consists of a series of stages in sequence. So, if we simply take our system that we are developing to a group of end users and ask for their feedback that is tell us what you think about the system, tell us whether you find this system usable or not, you may think that is one way of evaluating usability, but it is not as simple as that.

The feedback provided by the end users once they are asked this question may not be reliable. If we ask them the same question tomorrow, their answer may vary. Now, this variation may come due to the external factors such as their mental state, the environment in which they are working and similar such things. So, what we need is some sort of evaluation which will give us reliable conclusion.

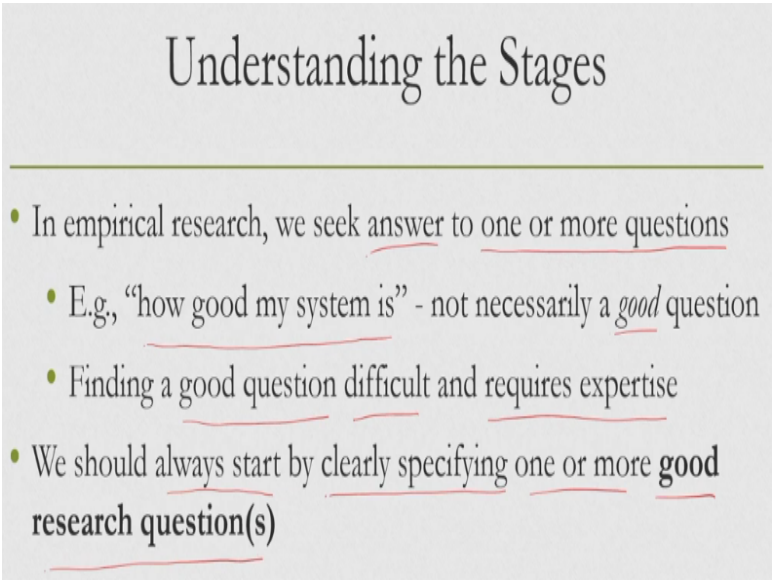
So the evaluation process should be able to produce reliable data from which we can draw reliable conclusions that is not an easy thing to do. It requires a very systematic and rigorous approach consisting of several stages which we are going to learn in this and subsequent lectures. So, what are the stages for empirical research?

(Refer Slide Time: 20:54)



Broadly there are four stages. First stage is identification of research questions. Second stage is determinism variables. Third stage is design of experiment and the fourth and final stage is analysis of empirical data. So, together these four stages constitute the overall empirical research method. We are going to learn about each of these stages in details with examples for better understanding.

(Refer Slide Time: 21:28)

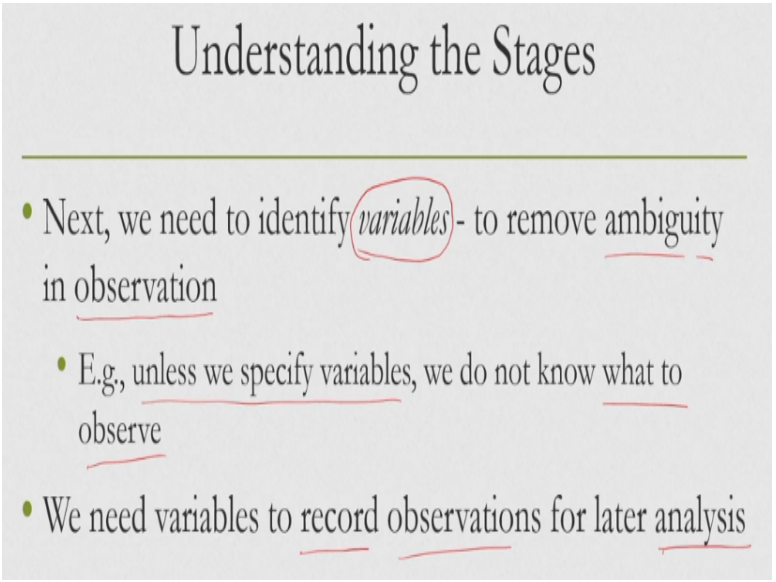


Let us have a brief understanding of these stages. Let us start with the first stage that is identification of research questions. So, in empirical research, we can think of this process as seeking answers to one or more questions. So, the whole purpose of empirical research can be thought of as trying to seek answers to few questions. It can be one question, it can be generally more than one questions.

For example, how good is my system? But one issue with that is what kind of questions we should frame for which we are looking for the answer, is this question how good my system is **is** a good question to ask or to seek answer of it is not necessarily a good question, why we will see later. So, it is very important to find a good question for which we seek the answer. The good question actually guides us in conducting the subsequent stages of the empirical research.

So, unless we are able to find a good question, our research method may be faulty. So, finding a good question is generally difficult and it requires some sort of expertise. So, our objective is to always start by clearly specifying one or more good research questions that is the first thing that we should do and that is the first stage also.

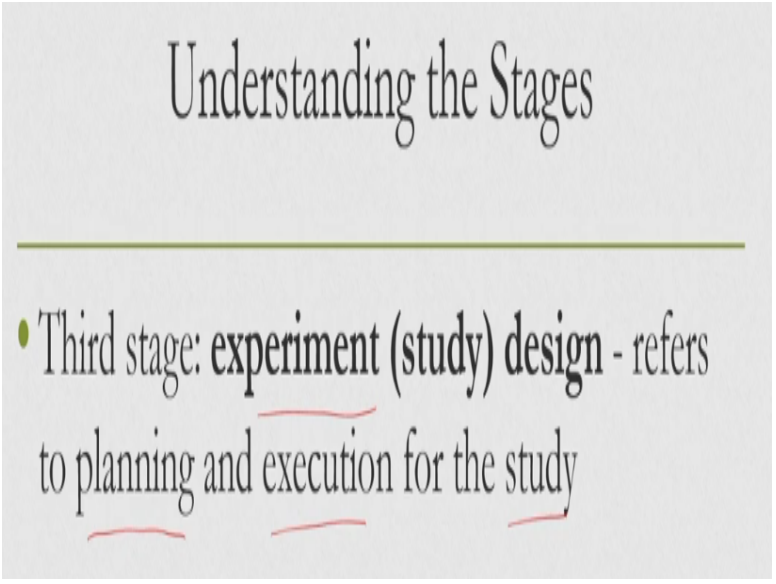
(Refer Slide Time: 23:01)



Next, what we need is to identify variables. We require to identify variables to remove ambiguity in observations. For example, unless we specify variables, we do not know what to observe. So, if we are unable to identify variables properly, then we will not be able to actually go for observations because it will not be clear to us what to observe and that will again lead to a faulty empirical research method.

So, after we formulate a good research question or more than one good research questions, we should go for identification of appropriate variables and the variables are required to record our observations for later analysis. So, that is the purpose of identification of variables to record our observation, unless we are able to identify the variables we will not be able to know what to observe and how to record it.

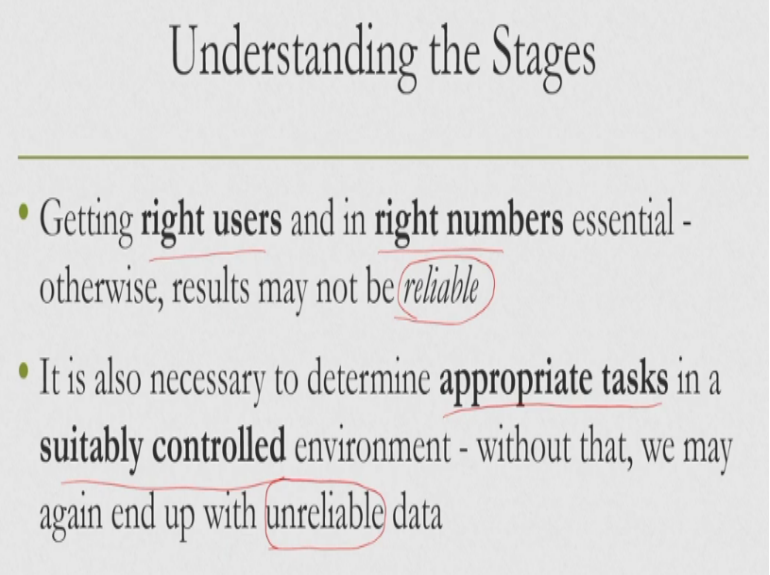
(Refer Slide Time: 24:07)



In the third stage, we need to design the experiment or the study. So, study not simply going to a group of end users and asking them questions, tell us this, tell us that. It is a systematic process, a very rigorous scientific process and we need to put in some effort to design the process, design the experiment. So, the experiment design stage essentially refers to planning and execution for the study.

So, the study is the core component of any empirical research method where we need to observe record observations which we will later analyse. Now, this observation and recording of observations requires some planning and some design of experiments that is the third stage of the empirical research method.

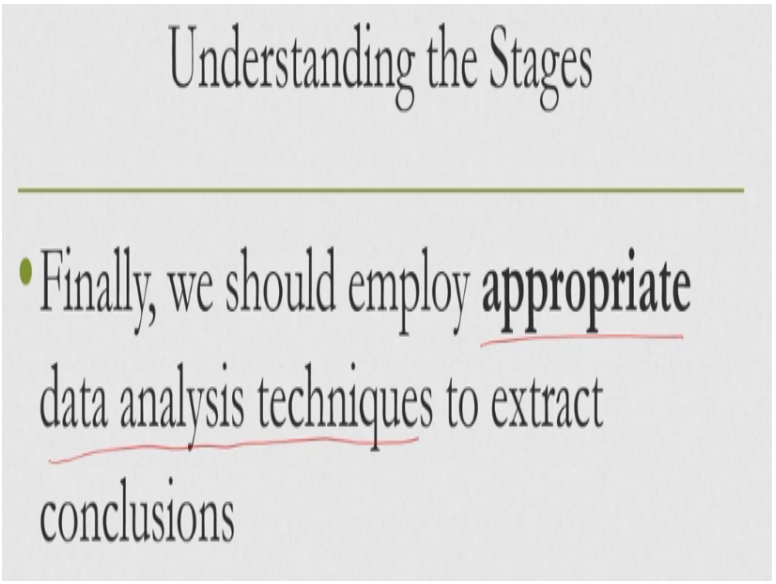
(Refer Slide Time: 25:00)



In fact, this is the core stage also, then comes the final stage that is analysis of the observed data. So, before we go to that one crucial point to note is why we need to design experiment is that we need to get right users and in right numbers to perform the experiment properly, otherwise the results may not be reliable. It is also necessary to determine appropriate tasks in a suitably controlled environment ad without that we may again end up with unreliable data.

So, even if we have good research questions and properly identified variables, still we cannot immediately start our experiment that is the controlled experiment unless we are able to find out right users in right numbers and able to identify appropriate tasks and able to control our environment suitably. If we do not do all these things, then we may end up with unreliable data leading to unreliable conclusion. So, these are very important and these requires us to properly plan the study.

(Refer Slide Time: 26:28)

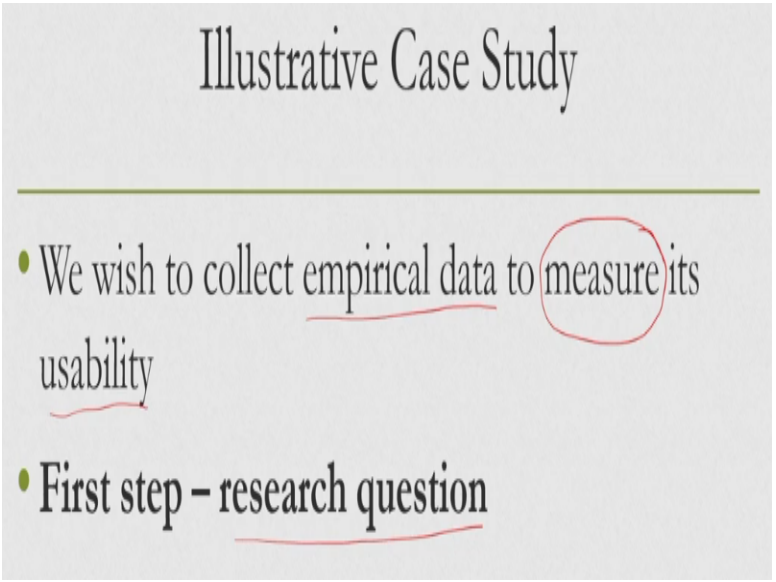


So, once that study is done and we are able to capture the data, the final stage is to employ appropriate data analysis technique to extract conclusions. Let me give you a quick example. Suppose, in an experiment we found that majority of the participants in the experiment that is majority of the end users who took part in the study opined that our system that we are testing has no usability and we directly conclude from here that our system is free from usability issues, not necessarily so.

We have to ensure that the opinion or the results that we obtained is reliable. For that we need to employ proper analysis techniques to first judge the reliability of the conclusion and then we can go for conclusion. So, appropriate analysis technique is also very important, which is the final and fourth stage of the empirical research method. So, in brief these are the four stages which we go through while performing empirical research to study about usability of our interactive system.

Now, let us go through each of these stages in details, let us try to understand each of these stages in details with examples. So, we will start with the first stage that is formulation of research question. Let us try to understand this stage with one example scenario. This will give us a better way to understand the importance of this stage.

(Refer Slide Time: 27:59)

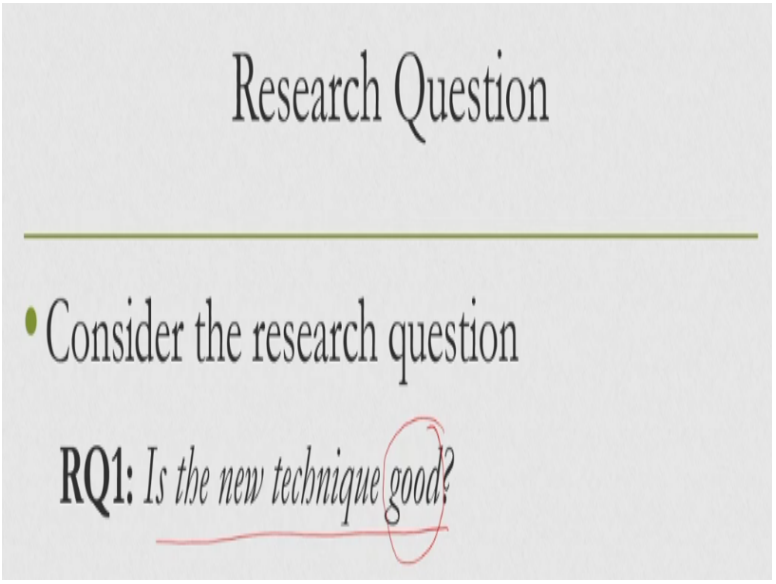


Suppose, we have developed a new text entry interface. We can think of it as a GUI or graphical user interface which can be used for text input. So, you want to input text and we have developed a GUI to provide the input to the computer. Now, that is one system that we have come up with. So, we have gone through all the stages, requirement analysis stage, design, prototype, evaluate cycle.

Then system design, stage coding stage, implementation stage, testing stage, testing of the code and finally we are left with the implemented and code tested system. Now, we want to understand its usability through empirical research. So, the first stage is we need to frame a research question for which we want to seek an answer. With that research question we wish to collect empirical data to measure usability or rather to get an idea of the usability measure.

So, the first step is we frame a research question. So, what can be a research question to measure usability of this particular interactive system that is text input interface.

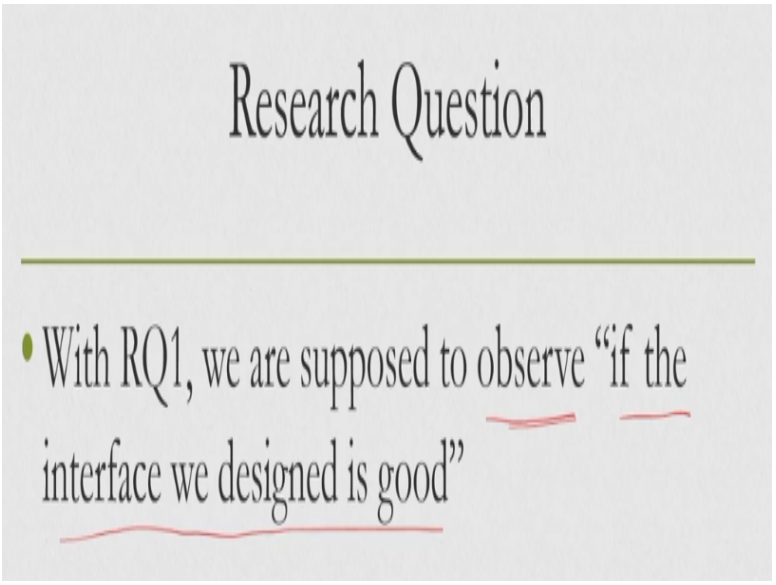
(Refer Slide Time: 29:30)



Let us start with a simple question straightforward, obvious, intuitive question. Is the new technique that we have developed good rather is the new system that we have developed good? Whenever you are asked to measure usability, many of us may immediately come up with such type of questions and we must try to seek answer to these questions through empirical research to know about usability of the product.

So, what do you think of this question, is the product good? It may appear that this is a good question because ultimately what we are interested in is to know whether our product is good, but is it really a good question? Let us try to understand it in a little more details.

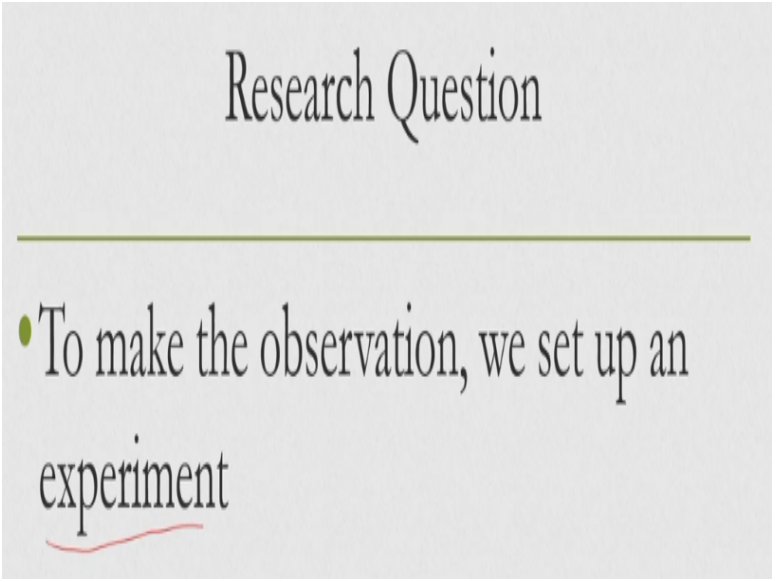
(Refer Slide Time: 30:29)



Now with this question, we are supposed to observe, note that empirical study is all about observations, so we are supposed to observe if the interface that we have developed is good.

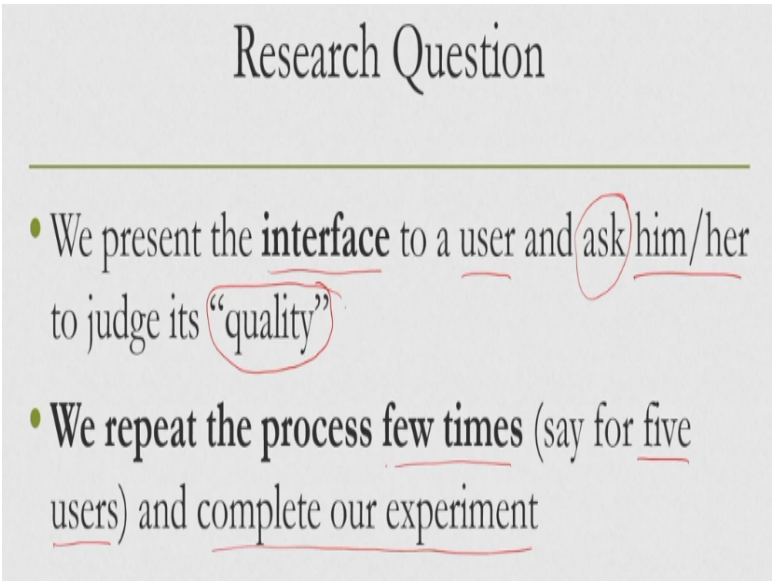
So, here when we use the term interface essentially it refers to the system and we are trying to observe whether the system that we have developed is good or not.

(Refer Slide Time: 30:57)



To make the observations, we set up an experiment. Now, at this stage we will not talk about the other stages of empirical study that is how to set up experiment, how to design experiment, how to identify variables, rather we will simply tell about the experiment. So, what is the experiment?

(Refer Slide Time: 31:17)

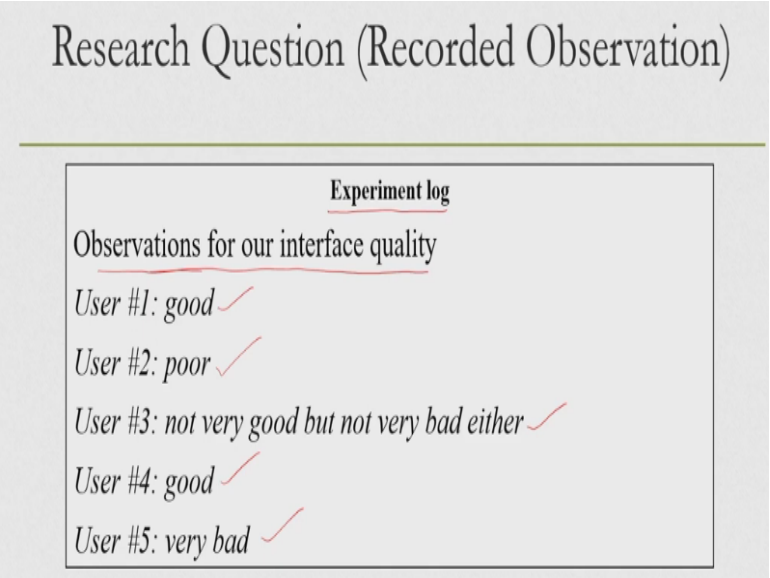


We present the interface to a user and ask the user to judge its quality. Note that we are interested to know whether the system is good, quote unquote good. So, it is essentially a qualitative attribute, so we ask the user to judge its quality. And since one user's data may not

be sufficient, so we repeat the process a few times. Let us assume that for five users we repeated the process and that is all about our experiment.

So, in the experiment what we did? We employed five end users, we give the system to each of them to learn about the system and then we asked each user to say something about the quality of the product because whether the product is good or not is a qualitative attribute.

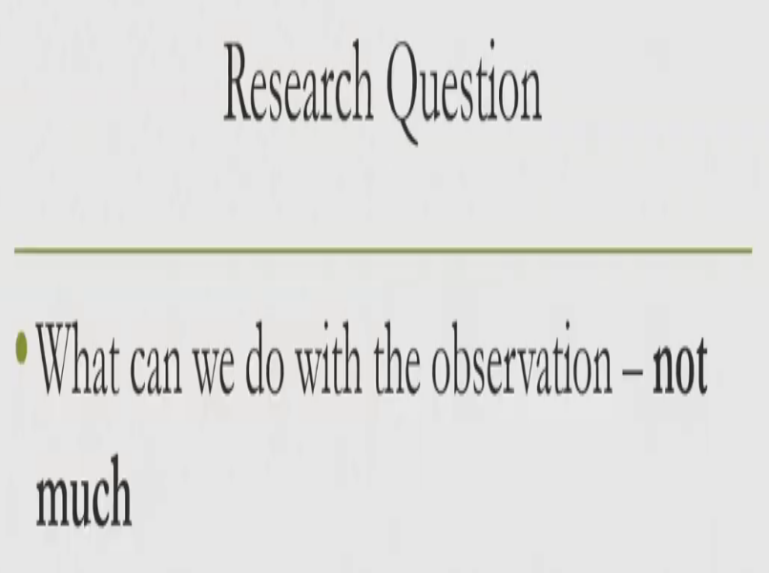
(Refer Slide Time: 32:22)



Let us assume that we got responses from all the five users. So, this is what the responses look like. So, we have logged the responses and produced an experiment log which records the observations for our product quality from the five users. So, user 1 mentioned that the quality is good, user 2 said quality is poor, user 3 said not very good, but not very bad either, user 4 said good and user 5 said very bad.

Now, why they gave these qualitative answers is not our concern, our concern is whatever responses they have given. So, there are five users each gave one response. User 1 said good, 2 said poor, 3 said not very good but not very bad either, user 4 said good and 5 said very bad that is our experimental log that means that is the observation that we made and recorded.

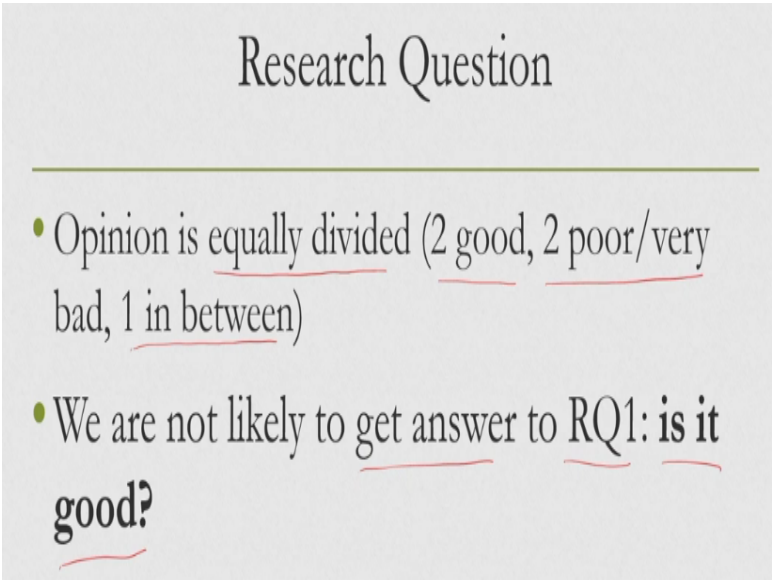
(Refer Slide Time: 33:35)



So, immediately the question that comes to our mind is now our observation is done, so we need to analyse and come to a conclusion, but the thing is what can we do with the observations? Can we really do anything with this observation? Obviously, we cannot do much. As you can see, there is no agreement between the users about the quality. So, two users, 1 and 4 said it is good, the user number 2 said it is bad whereas user number 5 said it is very bad.

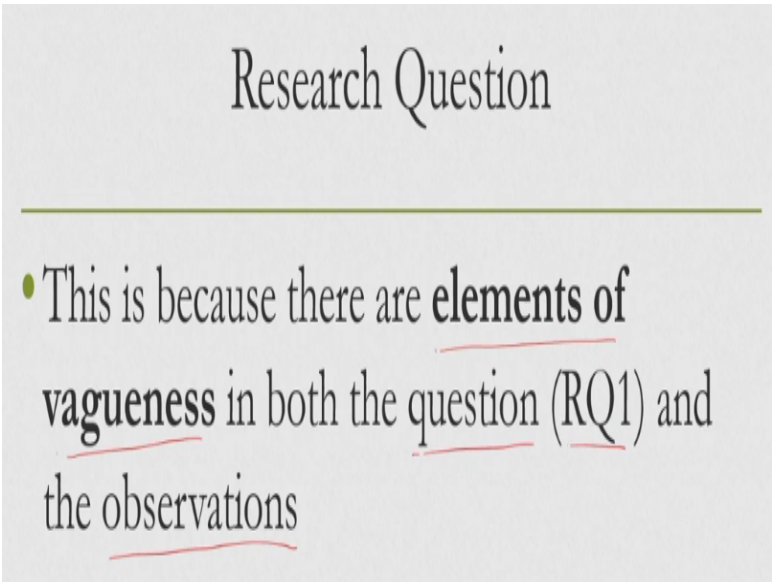
So, these two we can consider it to be referring to bad quality. So, two said good, two said bad. And user number 3 did not say anything definitive. So, according to that user it is not very good, but not very bad also, so a middle ground. Now, with this can we really conclude whether the interface is good or bad? Not really, we will not be able to come to such a conclusion.

(Refer Slide Time: 34:41)



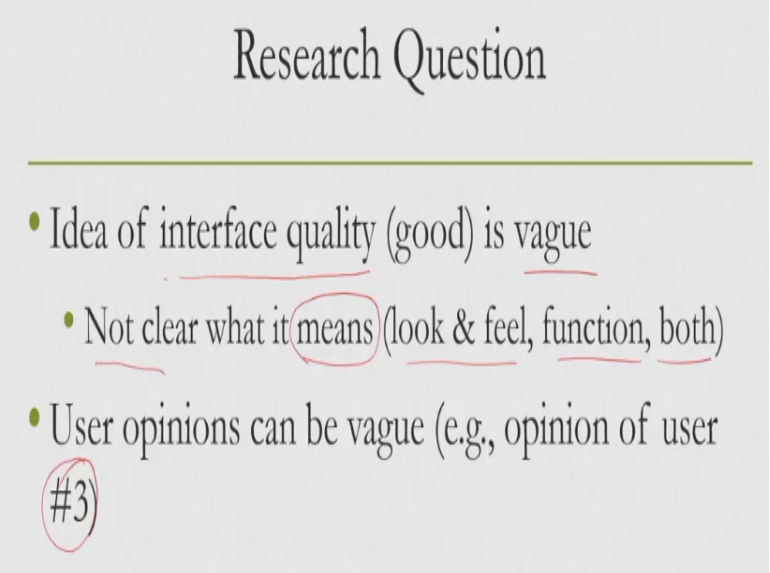
Why? Because the opinion is equally divided. Two said good, two said poor or bad and one is in between, neither good nor bad. So, we are not likely to get answer to the research question that we have framed that is whether our product is good. Note that in any empirical study, we try to seek answer to one or more questions. So, here in this simple case we framed one question whether our product is good. And from the observations, we really cannot get answer to this question because of the nature of responses that we have received.

(Refer Slide Time: 35:27)



Now, why we cannot get answer from the observations? This is because they are elements of vagueness in both the question that is the question that we framed as well as the observations. So, the question is not unambiguous, the observations are also not unambiguous.

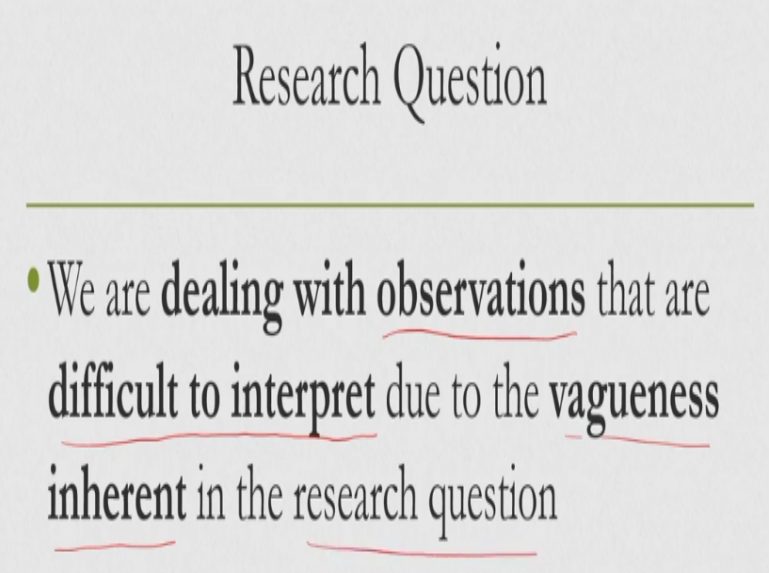
(Refer Slide Time: 35:59)



When we asked the users for their feedback on the quality whether it is good or bad, we actually did something which is not very clear to the user. So, the idea of interface quality whether it is good or bad is rather vague. It is not clear what it means. Does it mean that it is good in the sense of look and feel only or it is good in the sense of functions or it is good in the sense of both that is look and feel as well as functions? We did not specify or clarify on this issue while seeking feedback from the users.

The observations can also be vague as we have just noted. For example, the user opinion can be vague that is the opinion of user number 3, who said it is not very good, but not very bad either, a middle ground, so neither good nor bad in between. Now, if the question that is asked is vague, it is expected that the answers can also be vague, but that is not necessarily always correlated. So, even if the question is not vague, the answer can be vague.

(Refer Slide Time: 37:14)

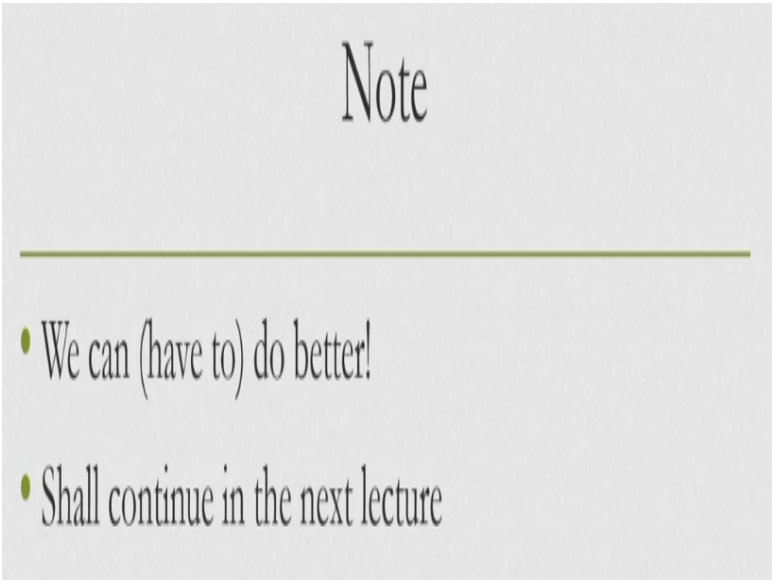


So, what is the problem here then? We are dealing with observations that are difficult to interpret due to the vagueness inherent in the research question. So, why this problem happened? Because we framed a question that is inherently vague. And when we asked the users to provide answers to this question or rather provide feedback on the basis of this question, the vagueness appeared in their responses as well.

So, the responses were difficult to interpret because of the inherent vagueness in the research question. So, then we have to somehow overcome this issue. So, we have to frame a question in a way such that the question does not have vagueness and accordingly we can expect the responses that are unambiguous, but that also depends on the kind of responses we are seeking. So, if we try to seek responses in the form of opinions, then they can be vague.

But if we try to seek responses that are not in the form of opinions, but rather in some other form then we can avoid vagueness. So, we have to improve on both aspects that is reduce or remove the vagueness in the research question as well as reduce or remove the possibility of vagueness in the opinions.

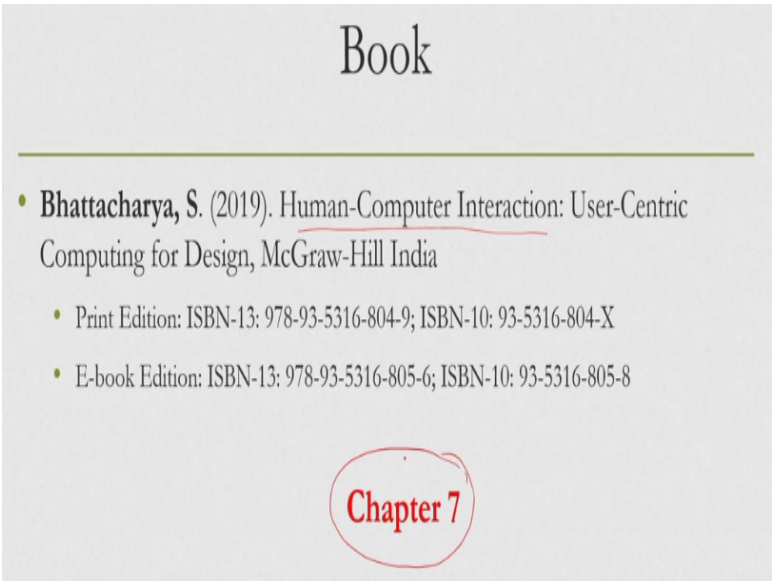
(Refer Slide Time: 38:39)



So, we can do better, in fact we have to do better, and how this betterment can be done we will learn about those in the next lecture. So, with that we have come to the end of this lecture. In this lecture, we have just introduced the first stage that is the formulation of appropriate research question which is the first stage of any empirical research method.

We have also learned the need for empirical research and the way it can be done that is the four stages which have to be performed in sequence. I hope you have understood the concepts that we covered today. We will continue our discussion on these concepts in next and subsequent lectures.

(Refer Slide Time: 39:24)



Whatever we are discussing can be found in this book Human-Computer Interaction: User-Centric Computing for Design. You may refer to chapter 7 to learn about the topics that

we are covering in this lecture and subsequent lectures. That is all for this lecture. Looking forward to meet you all in the next lecture. Thank you and goodbye.