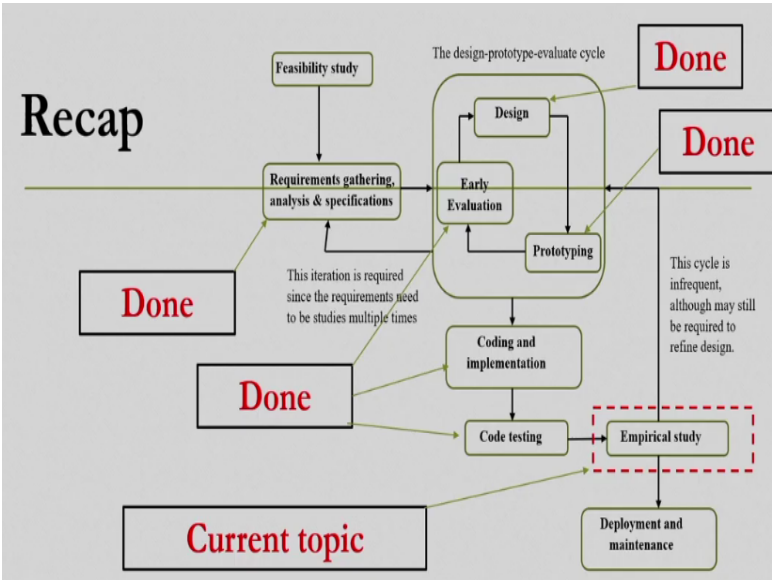
Design and Implementation of Human-Computer Interfaces
Dr. Samit Bhattacharya
Department of Computer Science and Engineering
Indian Institute of Technology – Guwahati

Lecture – 40
Empirical Usability Evaluation - 2

(Refer Slide Time: 00:56)



Hello and welcome to the NPTEL MOOCS course on design and implementation of human-computer interfaces. We are going to start lecture number 34 where we will continue our discussion on how to design experiment. So, before we start like we do for every lecture we will quickly recap what we have learned and where we are currently in the overall course and then we will go to the subject matter of this lecture.

So, we are discussing the interactive system development lifecycle containing several stages. We have covered most of the stages in the previous 33 lectures including case studies related to the outcome of the stages. So, let us quickly recap what are the stages that we have covered and what we are currently going to cover. So, we have covered the first stage requirement gathering analysis and specification. We are assuming this to be the first stage, although in principle feasibility studies should be the first stage.

However, as we mentioned earlier, we will ignore feasibility study in this discussion and assume that the lifecycle starts with the requirement gathering stage. So, we covered that in details. Next, we covered the design stage, design of interface and interaction. This was

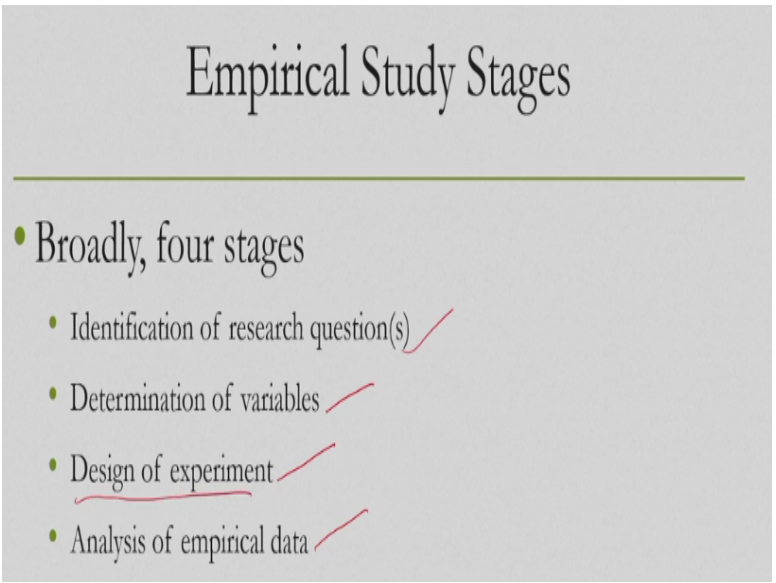
followed by the prototyping stage. How to create prototypes, followed by the evaluation of prototypes, how we can quickly evaluate prototypes, poor usability issues.

And this form the cycle which we need to perform probably many times before we finalize on a stable interface and interaction design. This was followed by design of the code or code design stage followed by the coding and implementation stage where we have learned about how to use good code practices for implementing the system. Next, we covered the code testing phase, there we learned about different ways to test our code to find out bugs in the code and to come up with an error free code.

Currently, we are discussing the next stage that is the empirical study stage. We have already covered basic concepts of the stage and we are going to continue our discussion on the empirical study in this lecture as well. Just to recollect empirical study in the context of interactive system design refers to the fact that in order to identify usability issues with our end product the interactive software after code testing, the testing does not stop, we need to test for usability.

Now, we do that in the empirical study stage where we set up controlled experiment, imply some end users, collect behavioural data of the end users while they use the developed system and we analyse the data to come to a conclusion about the usability of the system. Now, the empirical study broadly consists of four stages as we have mentioned in the earlier lectures.

(Refer Slide Time: 04:12)



Stage number 1 is identification of research questions, what do we do in this stage? In this stage, we first identify a suitable question for which we perform the study. So, a study means we are trying to seek answer to some question, those questions are called research questions. Unless we are able to identify suitable research questions, it will be difficult to proceed further in the empirical study.

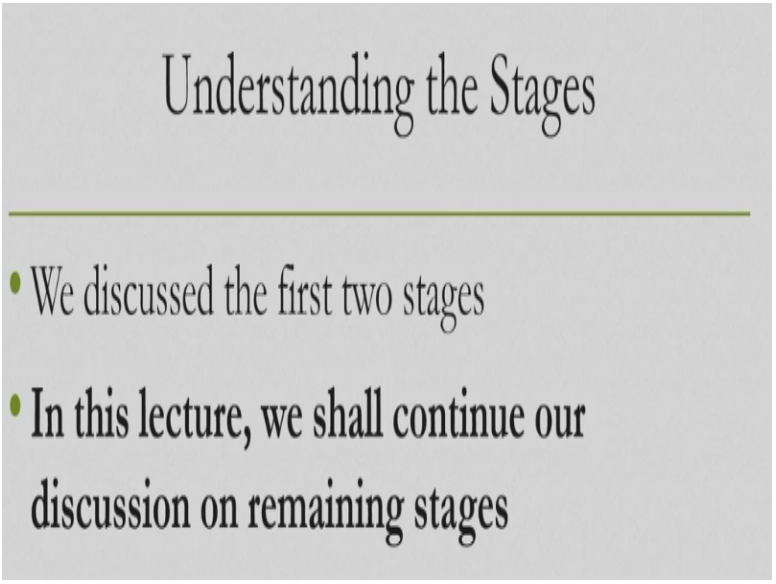
It will be very difficult to set up the experiments and collect data. So, the very first and essential step is to identify suitable research questions. Also, it may be noted that generally in literature you will find that instead of research questions in empirical study we try to deal with hypothesis. So, they are essentially the same concept, a research question leads to two hypothesis, null hypothesis and alternative hypothesis.

In the null hypothesis, we try to say something that is opposite to what we want to establish and an alternative hypothesis we say something that is what we want to establish. Our objective is to statistically refute or establish either of the hypothesis. So, either we refute the null hypothesis based on statistical evidence or we establish the alternative hypothesis based on statistical evidence that is how we proceed.

The next is determination of variables. This is also very important to let us know exactly what we are going to observe and measure. So, unless we are able to identify the variables, we cannot decide what to observe and how to record, so identification of variable is the other crucial component that we should take care of. Third is the design of experiments. So, once research questions or hypotheses are framed and we are able to identify the variables, we go for design of the experiment.

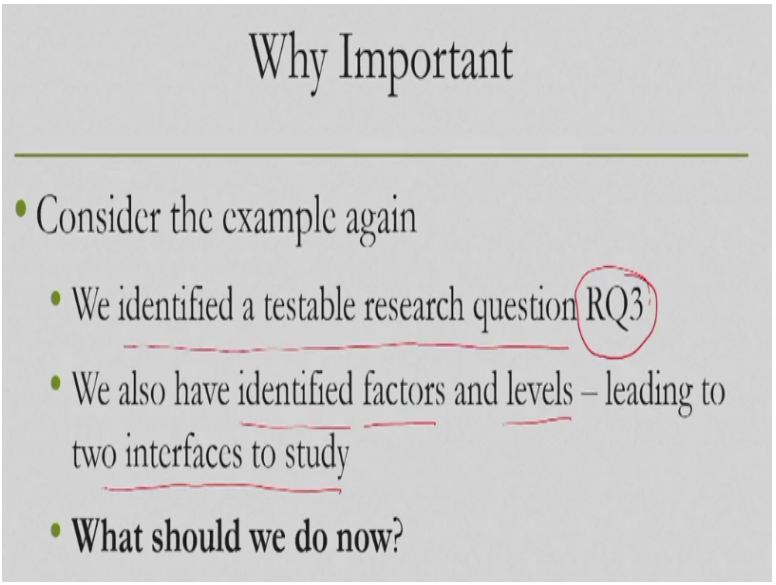
And finally, we analyse the observed data. In this lecture, we are going to discuss the third stage design of experiment.

(Refer Slide Time: 06:36)



So, in the earlier lectures we have already covered the first two stages that is identification of suitable research questions as well as identification of variables. In this lecture, we are going to concentrate on how to design the experiment, what are the issues involved, how to overcome those issues and how to come up with a suitable design of the experimental procedure that we are going to perform to collect behavioural data? So, we are going to talk about experiment design in this lecture.

(Refer Slide Time: 07:09)

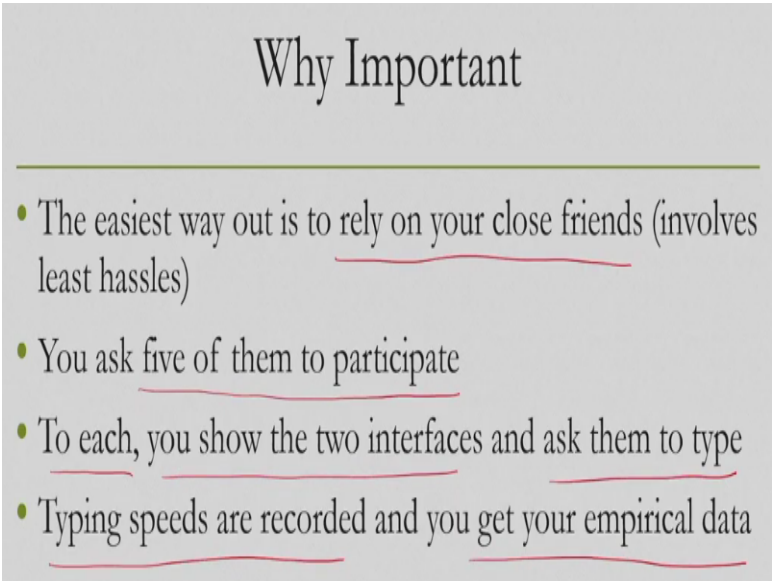


Let us try to understand why experiment design is important. Let us revert back to our earlier example, the research question 3. Now, we have already identified a testable research question RQ3. Remember, the question is testable if we can measure the dependent variables in a quantitative manner. We also have identified the factors or the independent variables and

the levels or the values of those variables for that research question that is the two types of interfaces that we considered, that is our design and existing design in the form of MS Word.

So, because there were two levels, so there were two interfaces that we used for our study. Now, the first two stages are done. So, we have identified the research question and we have identified the variables, what we should do next? Of course, our objective is to answer the question that which of the interfaces is faster? So, if this is asked of you that now try to find out the answer to this question, what you will do?

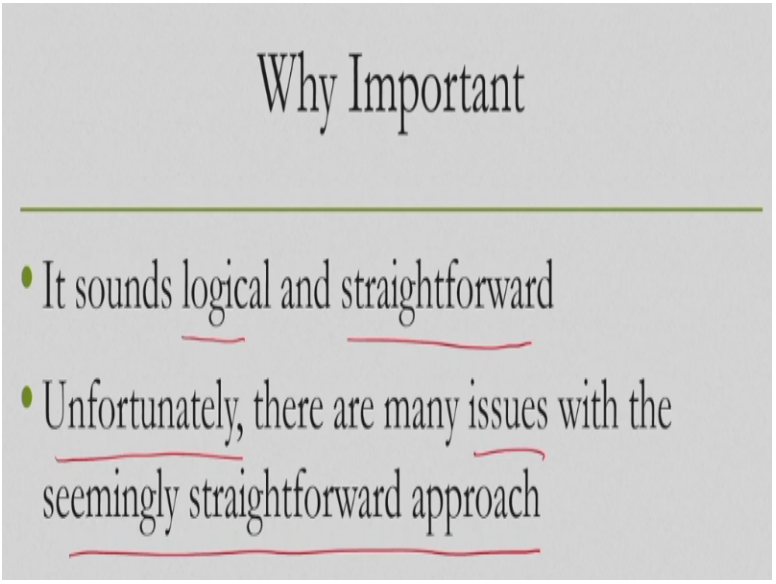
(Refer Slide Time: 08:42)



Most likely, we will follow the easiest way that is we will rely on our close friends to get participants who will participate in the study, who will be asked to perform the tasks and generate their typing speed data. So, of course we require participants and we approach our friends assuming that to be least problematic to get the data. So, you have asked 5 of your friends to participate.

To each of these friends you show the two interfaces and ask them to type, you have given them the typing tasks. Then you record it the typing speeds. And with that you got your empirical data based on which you can come to a conclusion. Sounds pretty simple, straightforward, and not much hassle. Of course, you have to somehow convince your friends to do these things which seemingly is not very difficult if you have good friends.

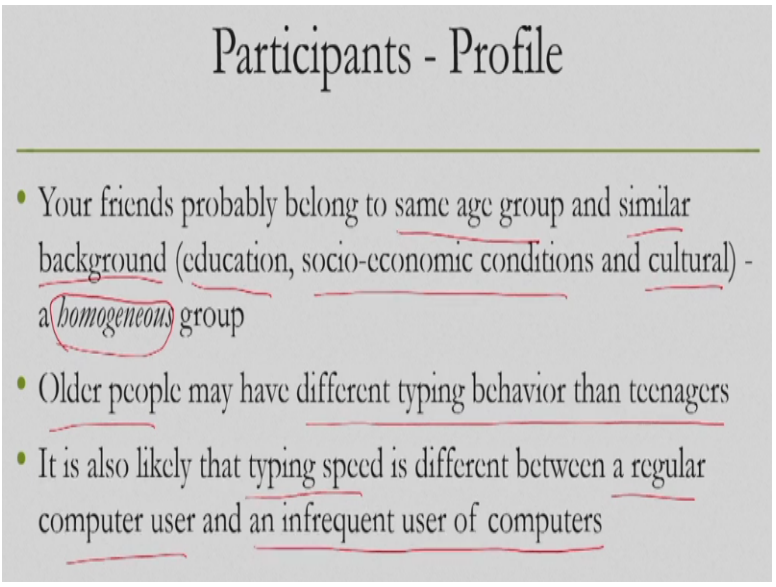
(Refer Slide Time: 09:58)



So, it sounds quite logical and straightforward, is not it? Unfortunately, there are many issues with this seemingly straightforward approach. Apparently, it may seem to you that you have followed all the stages. So, you have framed the research question, you have identified variables, then you have conducted the experiments by collecting participants that is your 5 friends and ask them to carry out the tasks and recorded their data.

And finally based on the data you analysed and come to a conclusion. So, apparently you have followed all the stages, which we mentioned earlier. Then can there be any issues? Was there anything wrong? There can be several issues with this seemingly straightforward approach, let us try to understand those issues. What can be those issues which are of concern to us?

(Refer Slide Time: 11:13)

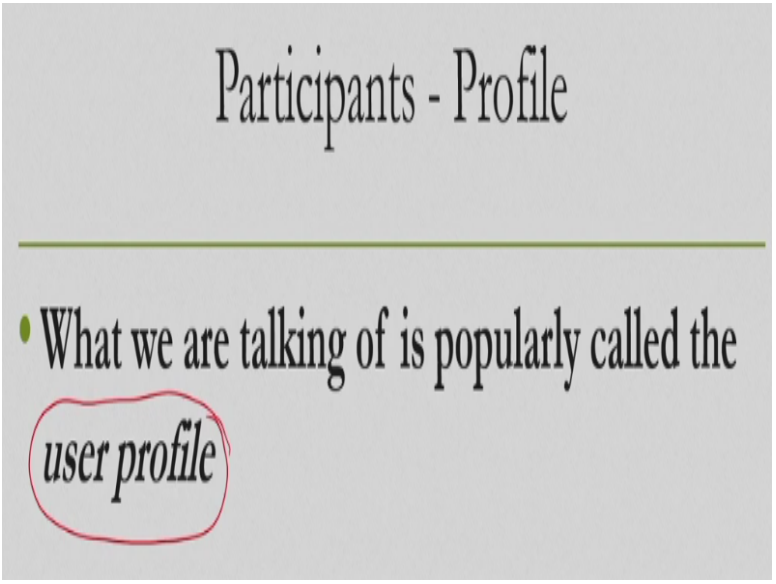


First thing is participants, let us see why choice of participants is very important. When you asked your friends to participate in the experiment and perform the tasks, you did not consider one thing. So, your friends are likely to belong to same age group with similar background, it is very rare that we choose friends who come from a very diverse background, generally our friends are of similar background in terms of education, socioeconomic conditions and cultural background.

Also, they are likely to be of same age group. So, we can usually term these people, these friends, the group of 5 friends to belong to a homogeneous user group that means a group having similar characteristics. Older people on the other hand may have different typing behaviour than teenagers. So, your friends assuming to belong to a particular age group, person from a different age group may behave differently which your friends will not be able to replicate.

For example, older people are likely to have different typing behaviour than teenagers. It is also likely that typing speed is different between a regular computer user and an infrequent user of computers that is another common sense knowledge that if your friends are regular computer users they are typing speeds are likely to be different than someone who is not a regular computer user. So, there are several issues with the choice of the friends as participants.

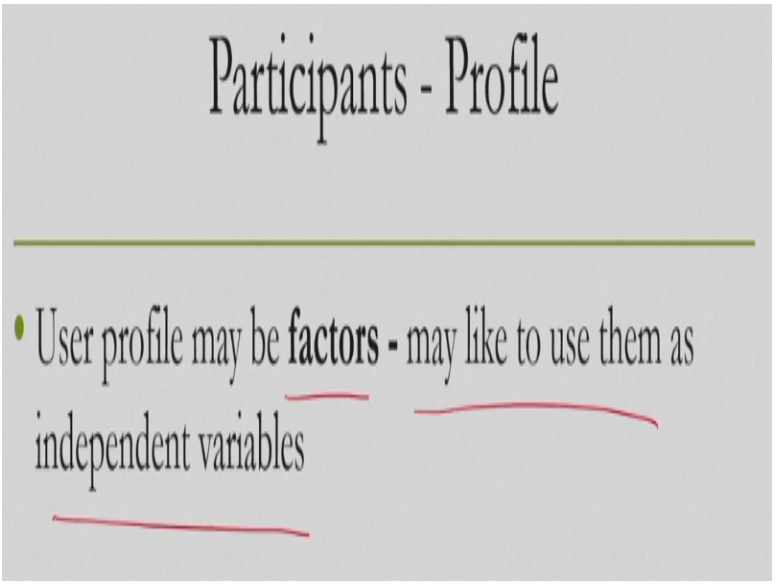
(Refer Slide Time: 13:01)



So, essentially, what we are referring to here is that we are talking of user profile. So, when we choose participants, we should be aware of the profile of the users. This is more popularly

called user profile and this term occur very frequently in the relevant literature. So, you should be aware of this term user profile. So, when we develop a system, we develop it for a group of users. As we have already mentioned earlier usability is applicable only for a specific group of users. Now, that group of users should have specific characteristics, which we call usually and popularly as user profile.

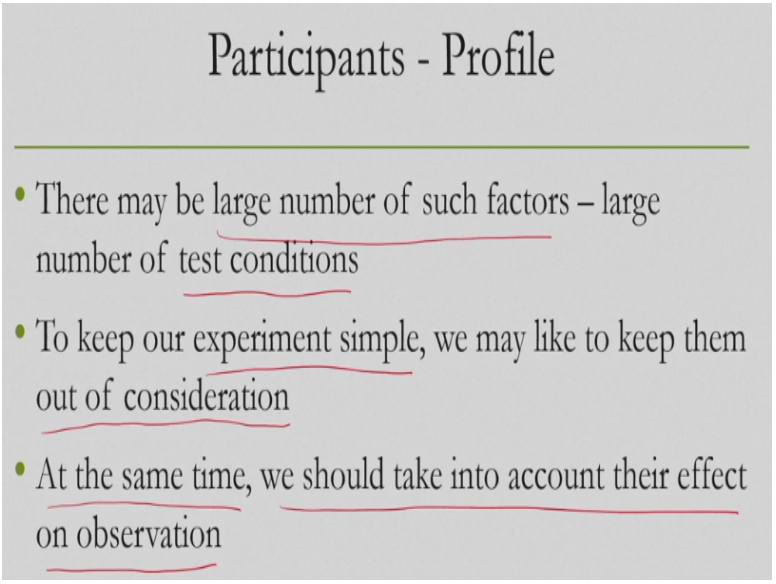
(Refer Slide Time: 13:47)



When we choose the participants for our experiment, definitely the profile of the participants should match with the profile of the intended user group of the system. Now the user profile, maybe factors that means independent variables. We may like to use them as independent variables or factors in our study. We may like to know the relationship between different parts of the profile, different characteristics in the profile and the outcome.

For example age, gender, economic background, cultural background, educational background, each of these can be a potential factor which can influence the outcome. So you may like to use them as factors as well.

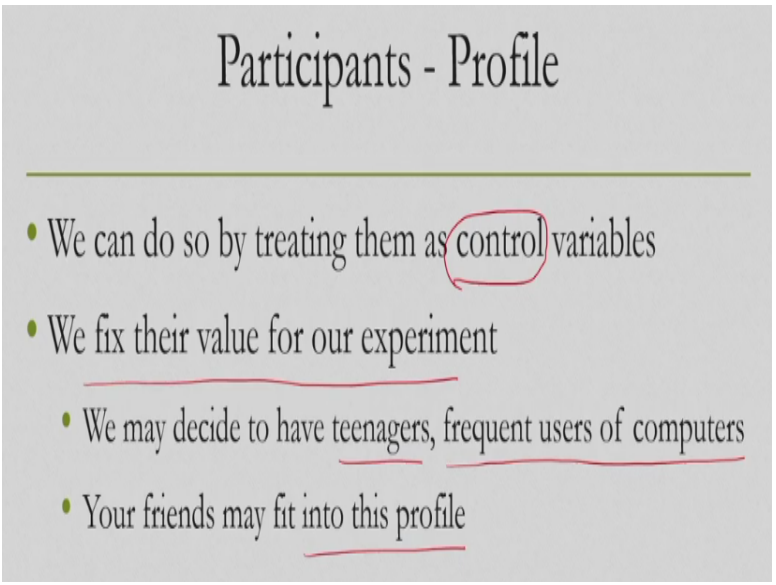
(Refer Slide Time: 14:40)



The problem is there can be a very large number of such factors which will in turn lead to a large number of test conditions. So, we have to carry out experiments for quite some time for a large number of test conditions, so that will increase the complexity of the experiments. In order to keep our experiments simple, what we can do is we may like to keep them out of consideration, keep those factors out of consideration.

That means we may not like to vary them, instead we may like to use them as control variables, we like to use them as constants throughout the duration of the experiment. So, we will fix their value. At the same time, we should take into account their effect on observation. Of course, we should be aware that these variables may affect the observation, so we should be very carefully controlling them.

(Refer Slide Time: 15:39)

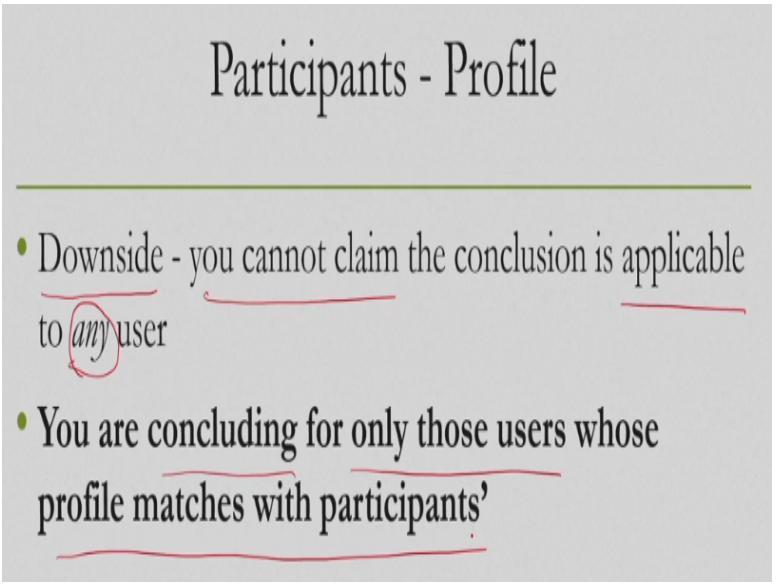


As I said, we can do so by treating them as control variables, earlier we have seen what is a control variable. These are variables which we use as constants that means we do not let them take different values instead during the experiment we fix the value for them so that it becomes a constant. So, the user profile also can be used as a control variable by fixing the profile characteristics.

For example, we may decide to have a specific user group that is teenagers, so we are fixing the age group. Also frequent users of computers, so we are fixing the experience. Suppose, we are fixing these two, so we are essentially using the two independent variables, namely the age group and the experience as controlled variables with fixed values. Now, if we do that, then we can roughly say that our friends may fit this profile.

So, in order to make your friends fit the profile, we probably need to analyse the situation a little bit more, try to identify the control variables, fix their values so that those matches with our friends' profile. So, then apparently it may seem that the problem gets resolved. So, we have chosen now a participant who belongs to the intended user group, although they do not represent the whole group, they represent a part of the group.

(Refer Slide Time: 16:59)



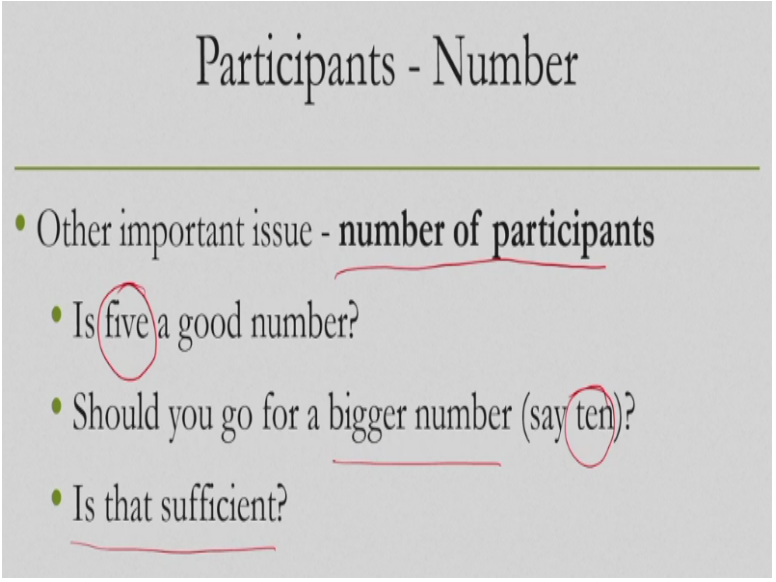
So, what is the downside? The major downside is that they do not represent the whole user group. So, if you are approaching your friends and asking them to be part of your experiment, then you cannot claim that the conclusions that you draw based on those experimental data is applicable to any user belonging to that intended user group. At most, what you can do? You

can say that you are concluding for only those users whose profile matches with those of the participants.

So, what we just discussed is that when we talk of a system it has a specific group of users. Those users are characterized by their profile. When we choose the participants to perform empirical study, we need to match the user profile with the participant's profile. If we do not match, then another way out is that we treat some of the profile characteristics as control variables and fix their values so that the participants' profile matches with the design that is fixed values for the control variables.

But in that case, whatever data we collect and conclusion we draw based on those collected data are not applicable to the entire user group. Those are applicable only to those users whose profile matches with the participants' profile. This is a very important thing that you should keep in mind while choosing your participants. That is about profile of the participants and whether those matches with the user's profile.

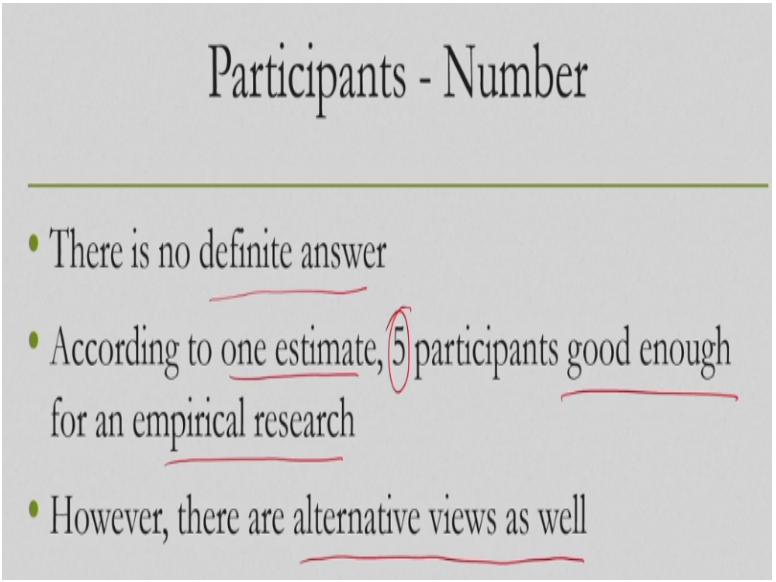
(Refer Slide Time: 19:05)



The other important issue that we should be aware of is number of participants. So earlier, we loosely say that let us approach 5 friends who will produce the data which we will use for analysis. But is that number 5 a good number of participants? Remember our objective is to come to a conclusion which is reliable. So, from the 5 users suppose we generated the data and we analysed that data to come to a conclusion. Can we say that those 5 users' data will lead us to a reliable conclusion.

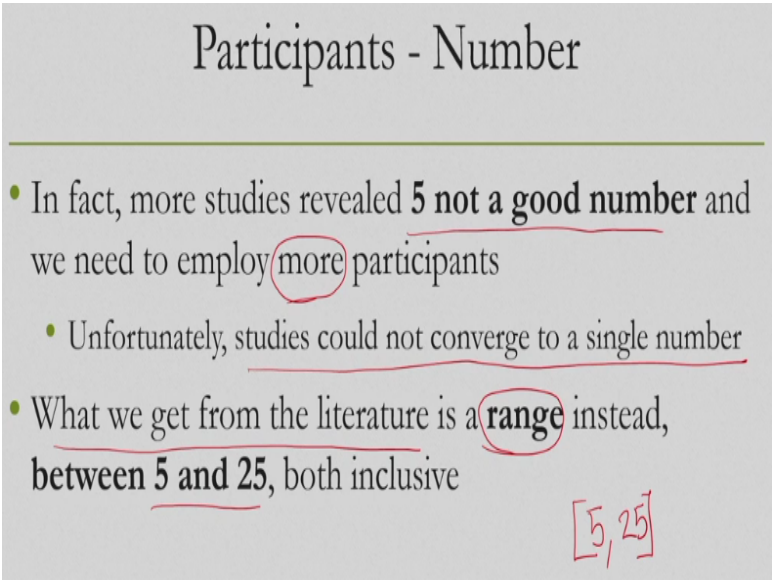
Later on, we will see that the data analysis that we perform on the observed data relies on statistical methods, so the number of participants and the corresponding amount of data that they produce should have some statistical significance. Now, what is a good number then? Should we go for a bigger number, say 10 to have the statistical significance of the data? Is that sufficient or we need even bigger value? These are some of the questions that we should be aware of and we should take care of.

(Refer Slide Time: 20:36)



Unfortunately, there is no definite answer to this question. We cannot say with certainty that so and so number of participants produces enough amount of data that can be used for reliable conclusion, so there is no definite answer to this question. According to one estimate, note that it is only estimate, 5 participants are good enough for empirical research, although there are alternative views as well; that means 5 participants are not good enough for empirical research.

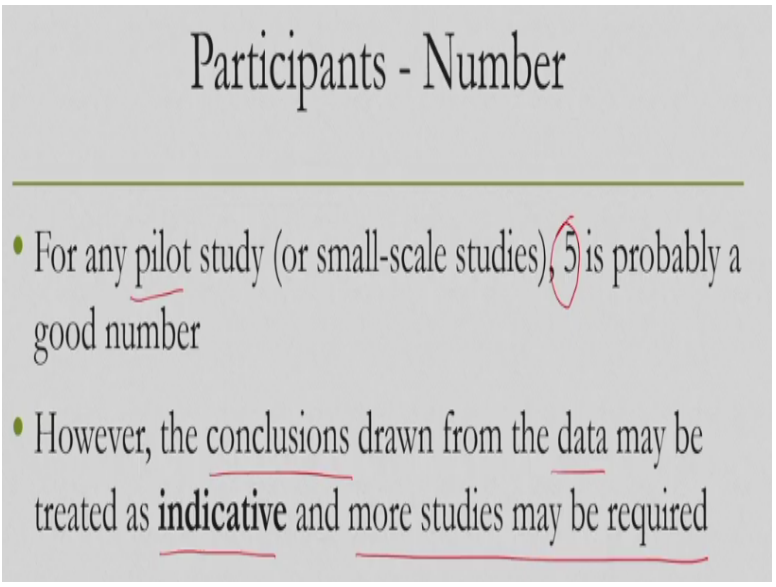
(Refer Slide Time: 21:18)



In fact, more studies have revealed that 5 is not a good number and we need to employ more participants. Again, unfortunately studies could not converge to a single number. So, although studies have revealed that data generated from 5 participants did not lead to a reliable conclusion, none of the studies could definitely say or agree on a single number. So, different studies tell us about different numbers.

So, what we get from the literature is a range of numbers instead of a single number between 5 and 25 both inclusive. So, essentially what it says is that the number of participants should belong to this range 5, 25; at least 5, at most 25 and both these numbers are included. So, we should have at least 5 participants or at most 25 participants to get reliable data for reliable conclusion.

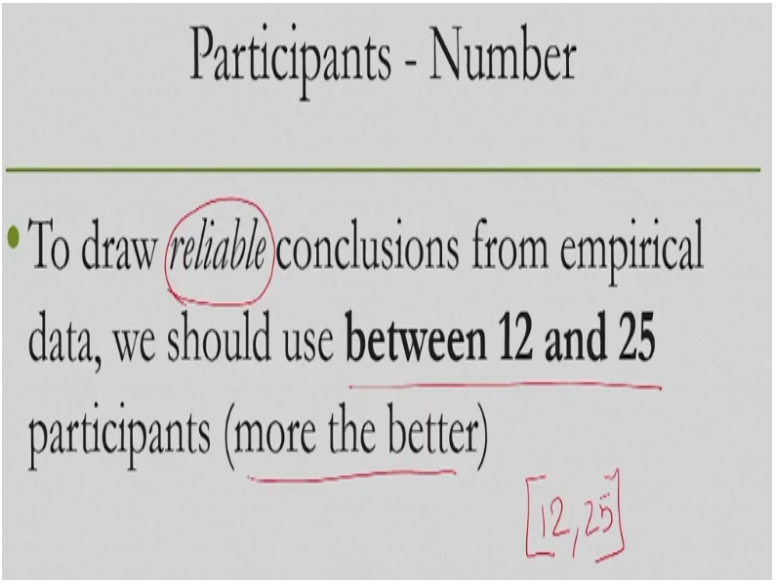
(Refer Slide Time: 22:37)



For any pilot study, that is small-scale studies that we have mentioned in the earlier lectures, 5 is probably a good number. So, if you want to conduct a pilot study or a small-scale study, then probably you can settle with 5 participants like asking your 5 friends. However, the conclusions that you can draw from the data from those 5 participants or less or nearby number of participants say 5 or 6 may be treated as indicative and more studies may be required.

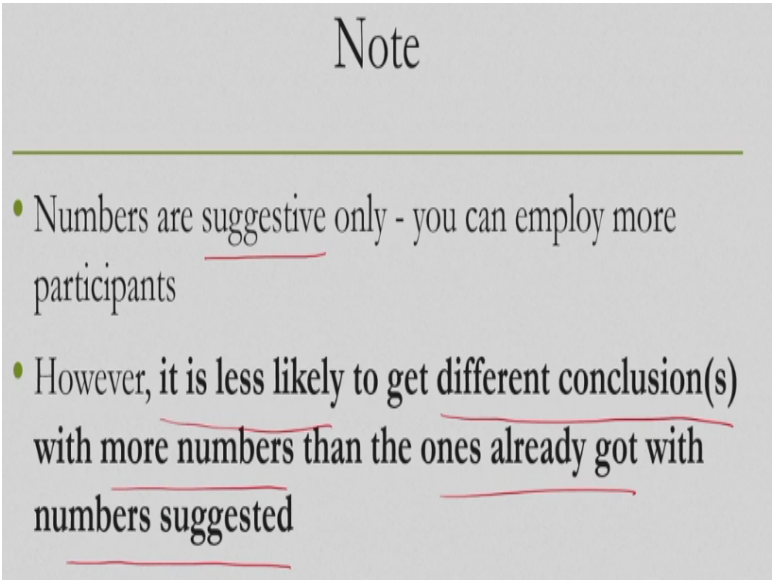
So, you cannot, ideally you should not conclude based on the data collected from 5 participants about the overall behaviour of the users. Instead, at most what you can say is that what is revealed by the data is indicative of the behaviour and more studies may be carried out to come to a definite conclusion.

(Refer Slide Time: 23:43)



To draw a reliable conclusion from empirical data, we should probably use between 12 and 25 participants both inclusive. That means, within this range of 12 at least and 25 at most. Although 12 is the lower limit, but more is better. So, that is what literature tells us that with 5 participants we may carry out pilot study for initial indicative conclusions and we may carry out more rigorous empirical study with between 12 to 25 participants, the more the better, that means closer to 25 probably is a good number.

(Refer Slide Time: 24:27)

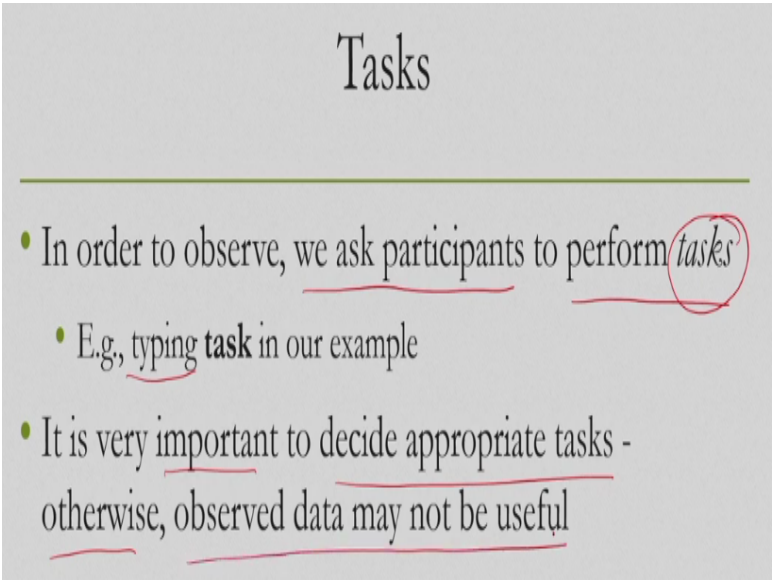


One thing you should note here is that these numbers are suggestive only. Based on some studies these numbers were found. It does not mean that you cannot employ more participants. So, if you wish to have more participants in your study, if you manage to get more participants, you can go ahead, no need to bother about this upper limit. However, it is less likely that you will get different conclusions with more numbers than the ones already got with the numbers suggested.

So, that means whatever conclusions you can draw based on the number of participants between 12 to 25 and whatever conclusions you are going to draw if you imply more numbers are not likely to be significantly different. So, you may not get a large number of significantly different conclusions from larger number of participants. So, essentially what it tells us is that between 12 to 25 is a good number, you can do most of the things with this number that lies within this range.

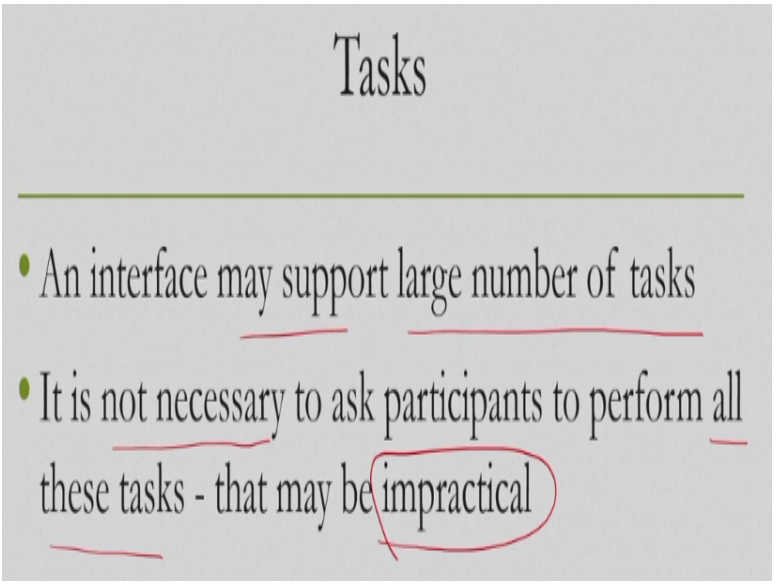
Of course, here it is suggested to have higher number of participants within this range that means around 25, but if you employ more participants then whatever conclusions you may get may not be significantly different than whatever you can draw with up to 25 participants. So, that is about answering two major issues if you want to go for experiment design that is how to choose the participant profile and how to choose the number of participants. Let us move our focus next to another issue for experiment design that is designer tasks.

(Refer Slide Time: 26:25)



In order to observe the user behaviour in our case, what do we do? We are supposed to ask participants to perform some tasks, some tasks that can be performed with the interface and interaction. In our example, the tasks can be typing of text strings. So, the typing task is one task that we asked our participants to perform to answer RQ3. So, it is very important to decide appropriate tasks before you go for the experiment, otherwise observed data may not be useful. So, unless you have carefully designed your tasks, the data that you may get may not be useful.

(Refer Slide Time: 27:19)



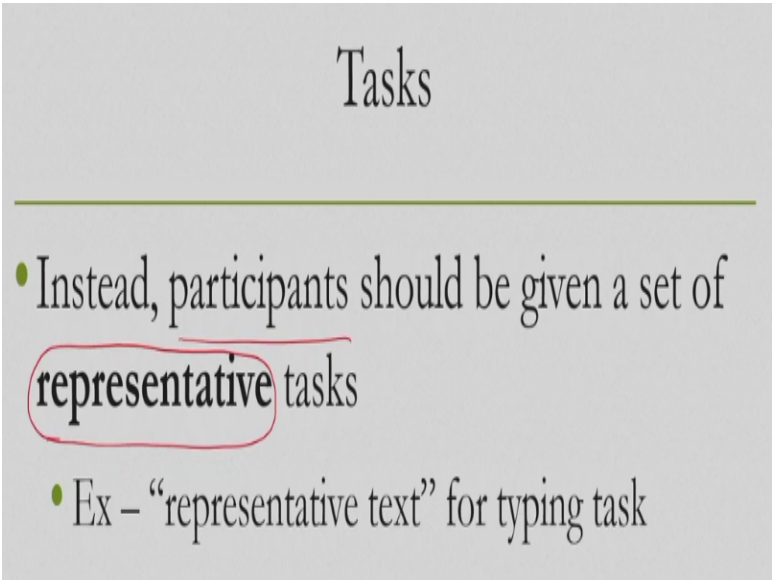
An interface may support a very large number of tasks depending on the number of features that are present on the interface. Now, it is typically not necessary to ask participants to perform all these tasks. In fact, an interface may support a very large number of tasks.

Consider the tasks that are supported by a text entry editor that is MS Word, enumerating all the tasks is a really humongous job and it will result in a very large number of tasks.

So, when we say that to carry out experiment, you first need to identify the tasks appropriately. Does that mean that whatever tasks are supported by the interface we have to identify and then accordingly we have to set up the experiment? Not necessarily so. Number of possible tasks with an interface may be very large, but we do not need to ask participants to carry out all the tasks, rather we do not need to design our tasks so that all these tasks can be carried out by the participants that may be impractical for many real-life systems.

As the number of tasks is very large, so it will take very large amount of time of the participants to complete the tasks, the participants may not like to do all these tasks for such a long time and they may quit in between. So, whole experimental process may become problematic to carry out.

(Refer Slide Time: 29:04)



So, instead of that, what you can do is you can ask participants to perform a set of representative tasks. Remember we introduced this term earlier where we said that a representative task is such a task which is expected to be performed by the users frequently in a real usage scenario. So, not all the tasks supported by an interface are done frequently by the users, few of those are frequently done and those few are the representative tasks.

So, our job is to identify those representative tasks and ask users to perform those representative tasks only. For an example, let us consider this typing situation. So we want to

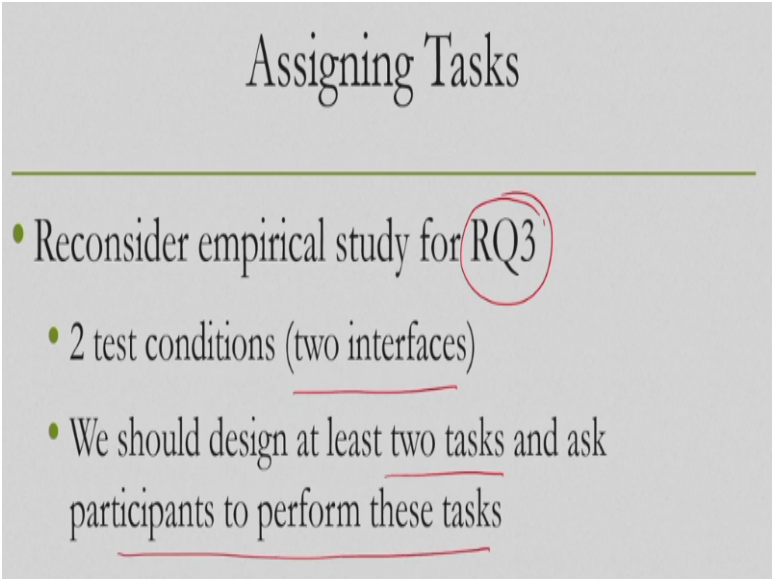
compare the two interfaces know which one is faster up, one is our design, one is MS Word design. So, here what can be representative task? There can be several tasks like save a file, open a file, type in a file, type on the screen, delete, add, etc.

But, how can we choose a representative task? Of course, when we are talking of text input tasks, the representative task is nothing but typing a text string, but what kinds of text string? Is it single character typing, a set of 5 characters that need to be typed? A full sentence to be typed? A full paragraph? Many possibilities are there, but what type of typing we should consider to be representative?

For that we can perform actually a statistical analysis technique on a corpus and find out a string that is commonly typed in real life situation for a particular language and that string we can ask the users to type during their typing session. So, during experiment we are going to give the user a specific string obtained through statistical method from a purpose of text which is likely to be typed frequently.

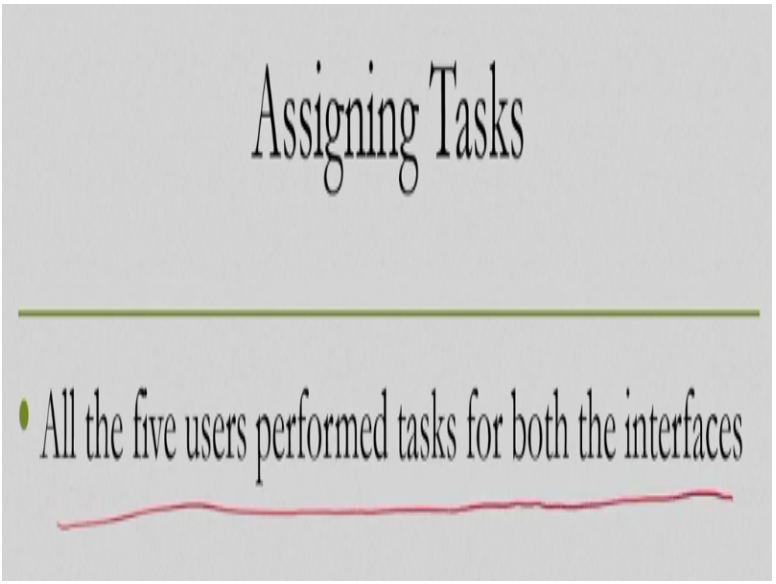
May not be the exact text but those characters and their relative positions with each other are likely to be typed by the users in real life situations that is a representative typing task for the particular experiment that we are using as an example.

(Refer Slide Time: 31:51)



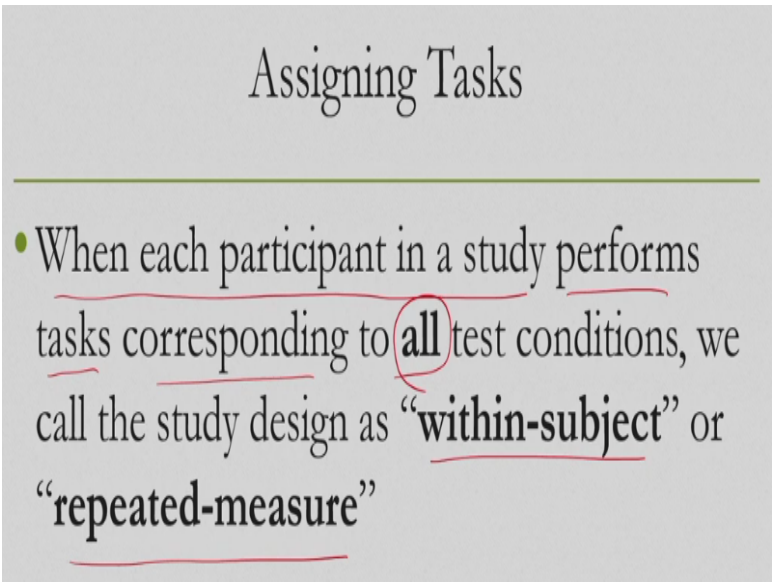
So, there were two test conditions for our empirical study related to RQ3 corresponding to two interfaces. So, we should design at least two tasks and ask participants to perform these tasks.

(Refer Slide Time: 32:07)



So, the next question is once we are able to identify the representative task, there are 5 participants, two interfaces. The next relevant question is how do we ask the users to type the tasks for those interfaces? All the 5 users are supposed to perform the tasks for both the interfaces.

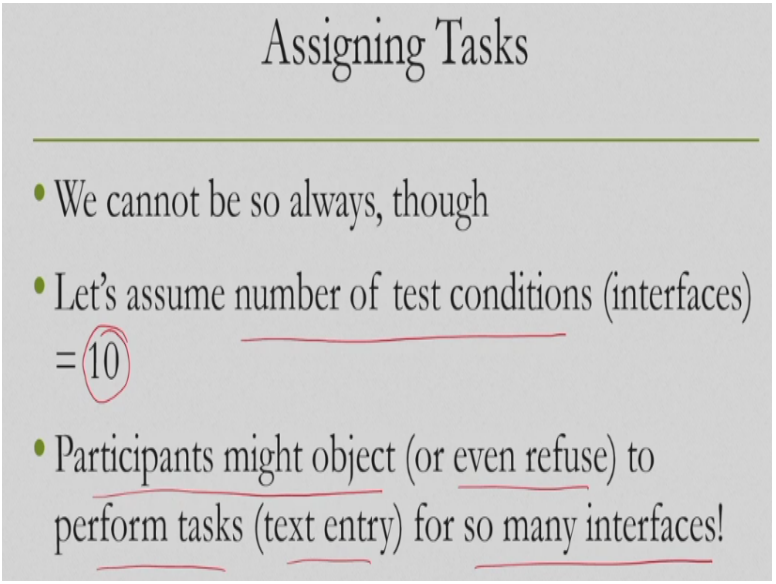
(Refer Slide Time: 32:29)



So, how do we assign these tasks? Before that we need to know about some terminology that is within-subject design or also known as repeated-measure design and between-subject design. So, what is within-subject design? When each participant in a study performs the tasks corresponding to all the test conditions, we call the study designers within-subject design or alternatively repeated-measure task design.

So, in our example we have two interfaces, 5 participants, each participant was asked to perform tasks on both the interfaces, so all participants perform the same way. So, our particular study design we can call repeated-measure or within-subject design.

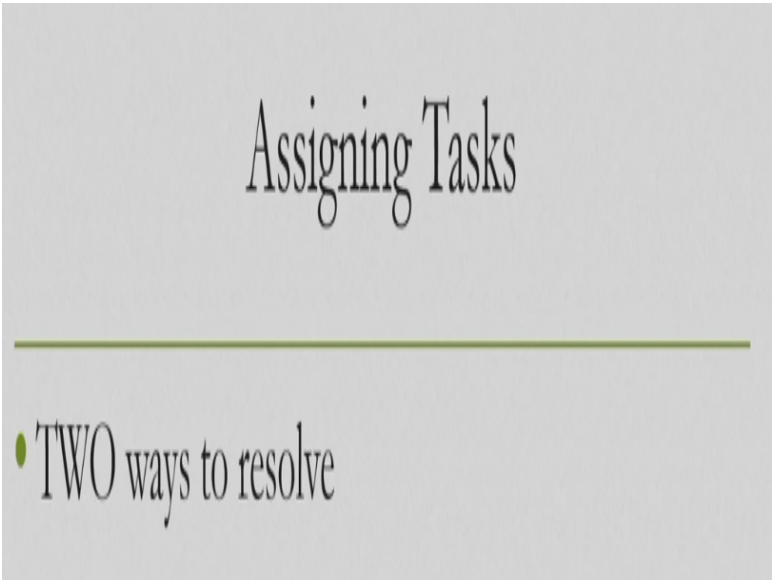
(Refer Slide Time: 33:27)



But that need not be the only way we can assign tasks. Sometimes there can be a problem. Let us assume that the number of test conditions that is the interfaces are 10. The number of test conditions is 10 that means there are 10 interfaces, so each participant has to type on 10 interfaces. Now, if we ask our participants to type on those 10 interfaces, then they may object, they may simply say that we do not want to type for so many interfaces.

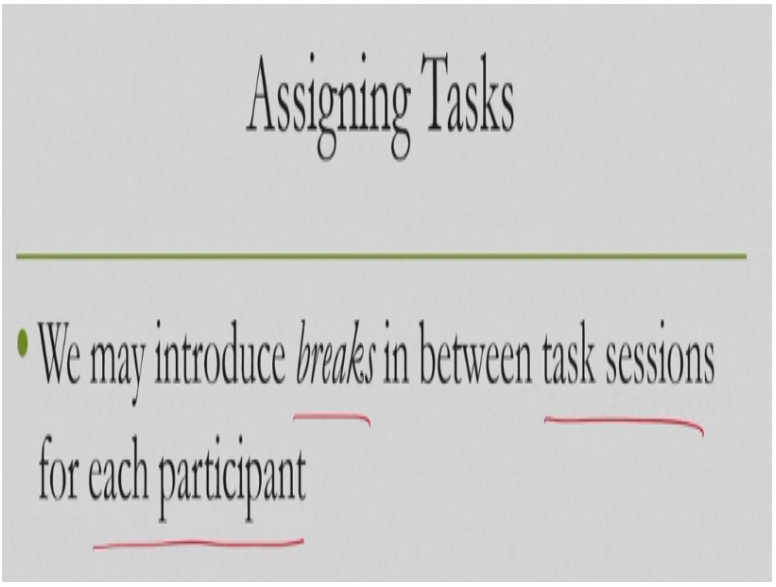
They may refuse to perform those tasks that is the text entry tasks for so many interfaces that is quite possible. Participants are after all human. So, if you ask them to do too many things, they may simply refuse to do that.

(Refer Slide Time: 34:23)



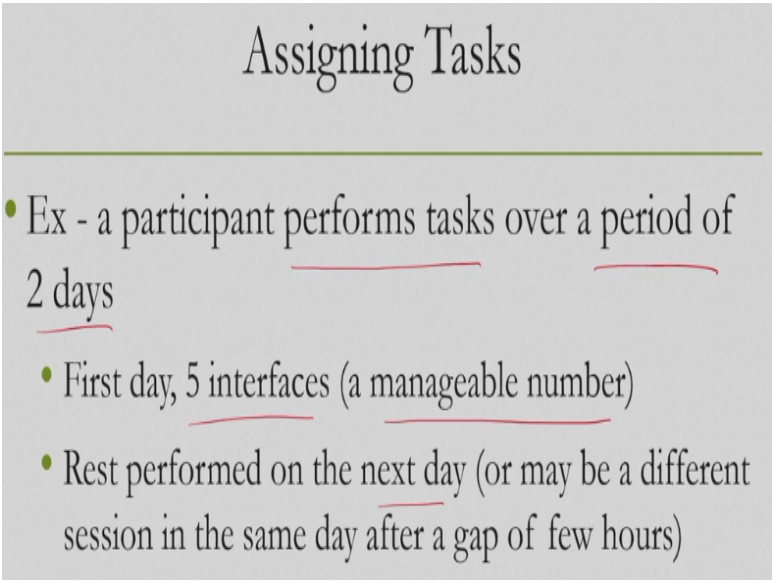
In that case, what we can do? There are two ways to resolve these issues.

(Refer Slide Time: 34:27)



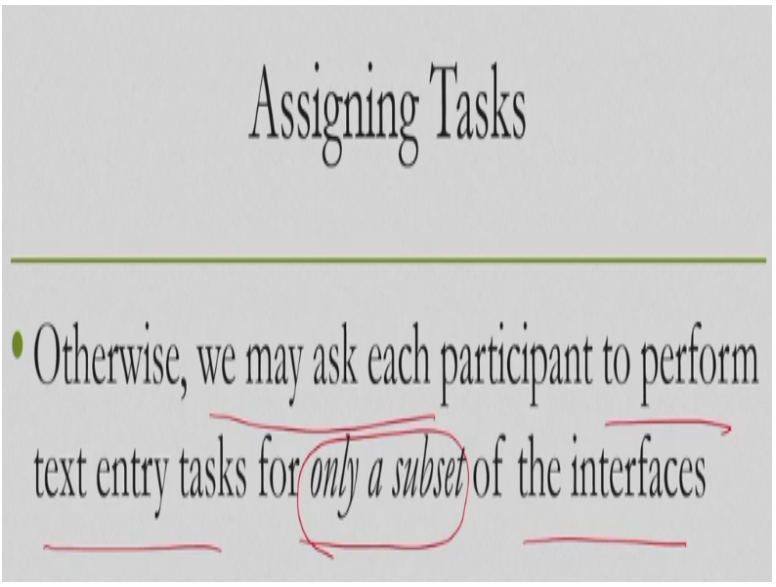
So, we may introduce breaks in between the task sessions for each participant. So maybe we can design the experiment in this way that there are several sessions for performing the tasks. In the first session, a participant performs with two interfaces. Then there is a gap, maybe 1 hour gap, then another session with two more interfaces, in that way total 5 sessions for 10 interfaces. So in that case, the participants may not feel the burden and the sessions can be spread over multiple days as well rather than on a single day that can be a straightforward solution.

(Refer Slide Time: 35:13)



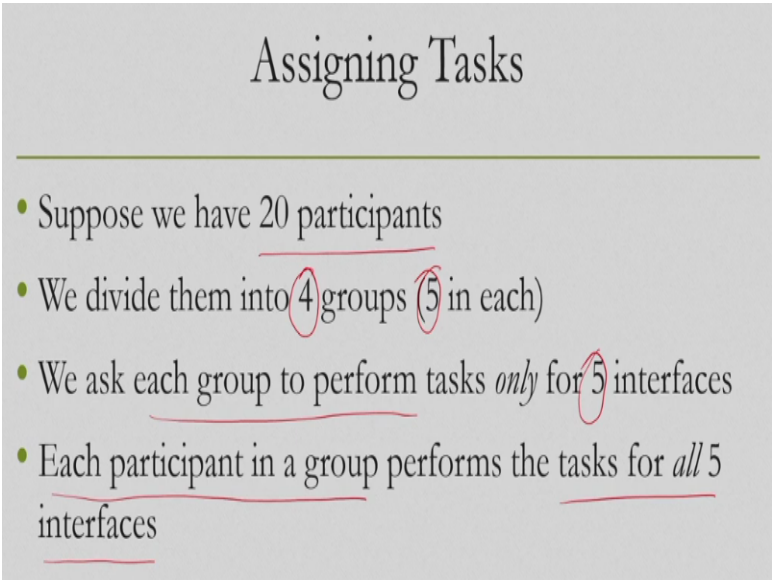
For example, a participant performs the tasks over a period of two days. First day, 5 interfaces, maybe over 3 sessions which is a manageable number. And the rest performed on the next day again over 3 sessions or maybe a different day with 1 day gap in between; anything is possible. So, these are manageable numbers and participants may not refuse in such cases.

(Refer Slide Time: 35:41)



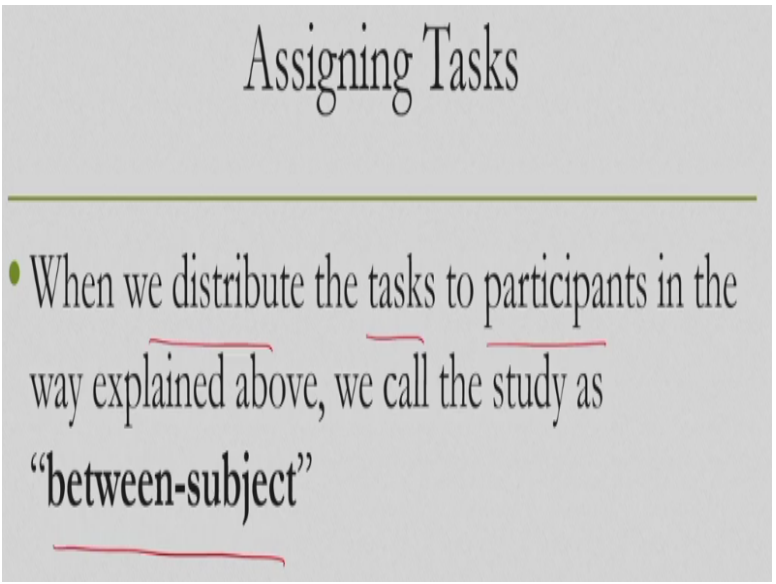
Otherwise, we may ask each participant to perform text entry tasks for only a subset of the interfaces. So, that is another way to do things. So, not all the participants are performing for all the interfaces, a participant is performing only for a subset of the interfaces.

(Refer Slide Time: 36:04)



Suppose we have 20 participants, we divide them into four groups. So earlier instead of 5, now we have 20 participants, we divided them into four groups, so 5 participants in each group. We ask each group of participants to perform tasks only for 5 interfaces. So, there are total 10 interfaces, so each group performs only for 5 interfaces. So, then each participant in a group performs the tasks for all the 5 interfaces.

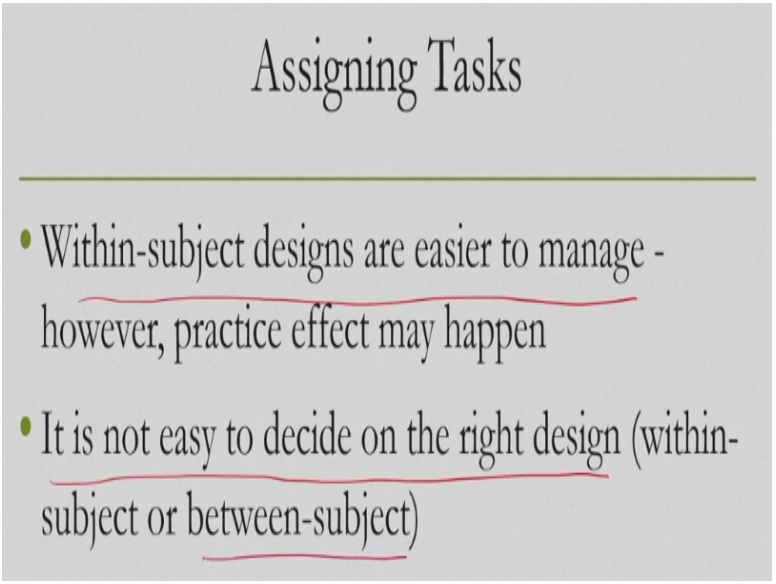
(Refer Slide Time: 36:42)



If we design our experiment in this way, that means we assign the tasks in this way then this is called a between-subjects study. That means when we distribute the tasks to the participants in such a way where not all participants participate in all the test conditions, then such a study is called between-subject study. So, we have within-subject or repeated-measure study design as well as between-subject study design.

These are related to how we distribute the tasks to the participants. If the numbers are manageable, then within-subject is fine, if the numbers are not manageable then between-subject can be used.

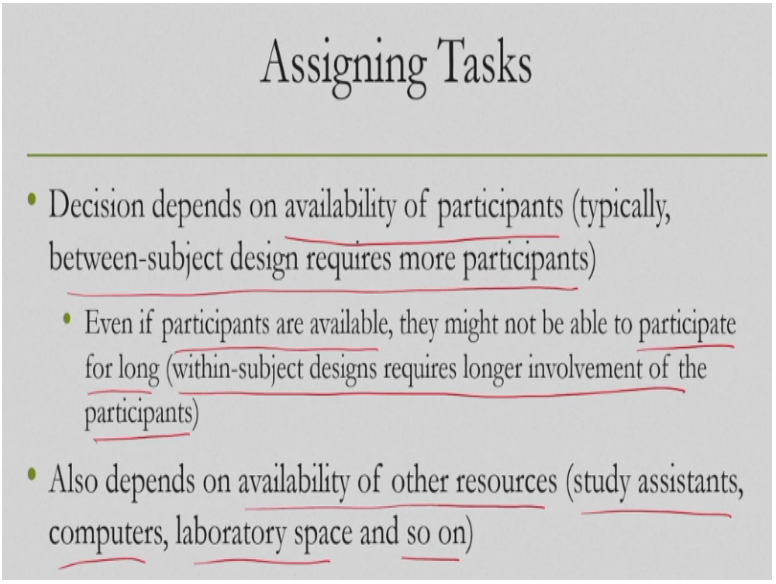
(Refer Slide Time: 37:23)



It may be obvious that within-subject designs are easier to manage. But the problem with within-subject design is that they are maybe the practice effect. So earlier we introduced the idea of practice if that is the sequence in which the tasks are performed by a participant or a user can induce some expectations or goals within the users' mind and because of that in subsequent task conditions, the outcome may get affected that is the practice effect as we discussed earlier in the previous lecture.

So, within-subject design although manageable, easier to manage, may lead to practice effect because all participants are taking part in all the test conditions. On the other hand, between-subject design reduces the practice effect, but then combining the data may be problematic because not all participants are participating in all the test conditions. So, both have positive as well as negative issues, pros and cons. So, it is not easy to decide on the right design, which one we should go for within-subject or between-subject.

(Refer Slide Time: 38:49)

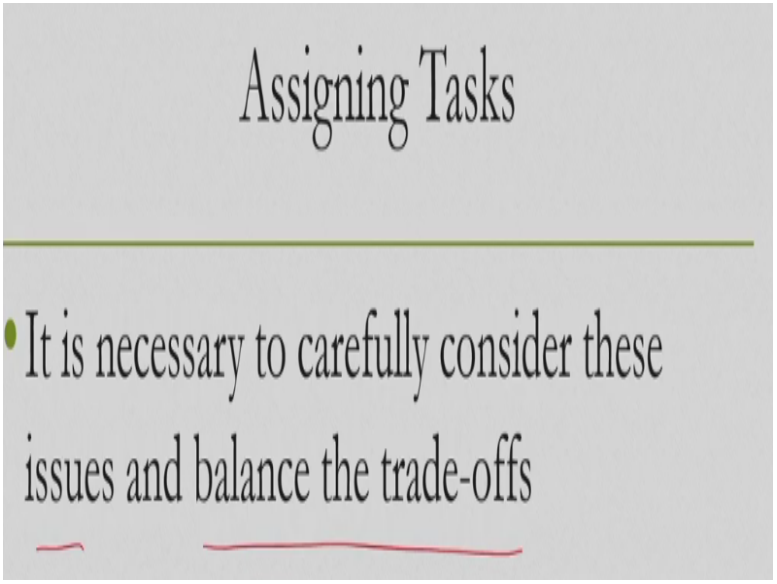


The decision depends on availability of participants. Typically, between-subject design requires more participants. So, if we have more participants and more test conditions, then we can go for between-subject design. Even if participants are available, they may not be able to participate for long that may be another condition, within-subject design requires longer involvement of the participants.

So even if we have less number of participants, participants may not be available for long maybe because of their professional commitments or physical abilities, they may not be able to participate for all the test conditions. So in that case, we have to go for between-subject design rather than within-subject design, although the number of participants may be less. It also depends on availability of other resources such as study assistants, computers, laboratory space and so on.

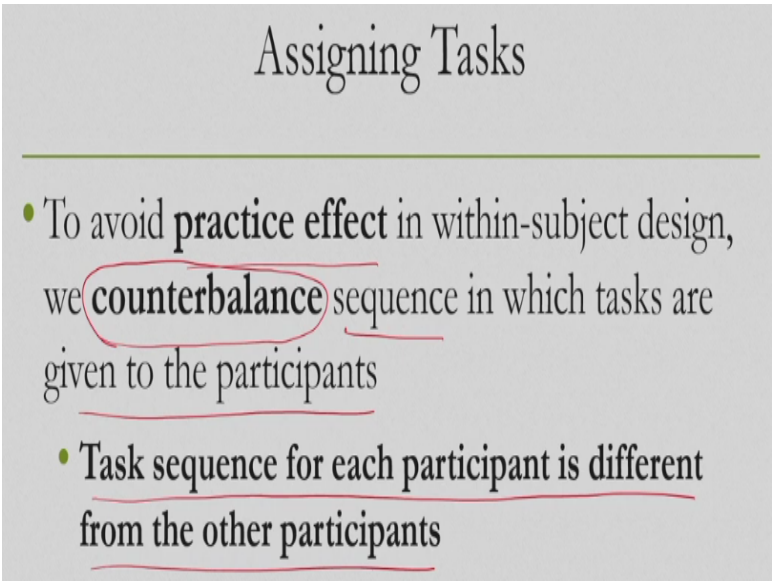
Definitely all these are supporting infrastructure that are required to carry out the empirical study. So, if they are available in sufficient quantity that determines what kind of study we can design.

(Refer Slide Time: 40:06)



So, it is necessary to carefully consider these issues and balance the tradeoff between the two design approaches, within-subject and between-subject.

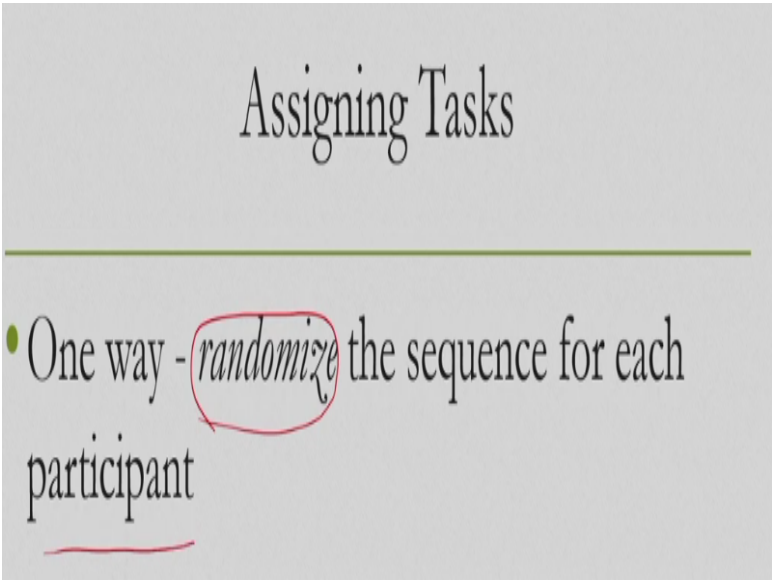
(Refer Slide Time: 40:17)



Now, if we go for within-subject then there is the issue of practice effect. So, to avoid the practice effect in within-subject design, we can employ one technique that is called counterbalancing. So, what we can do is we can counterbalance the sequence in which the tasks are given to the participants. So, task sequence for each participant is different from the other participants in that counterbalancing technique.

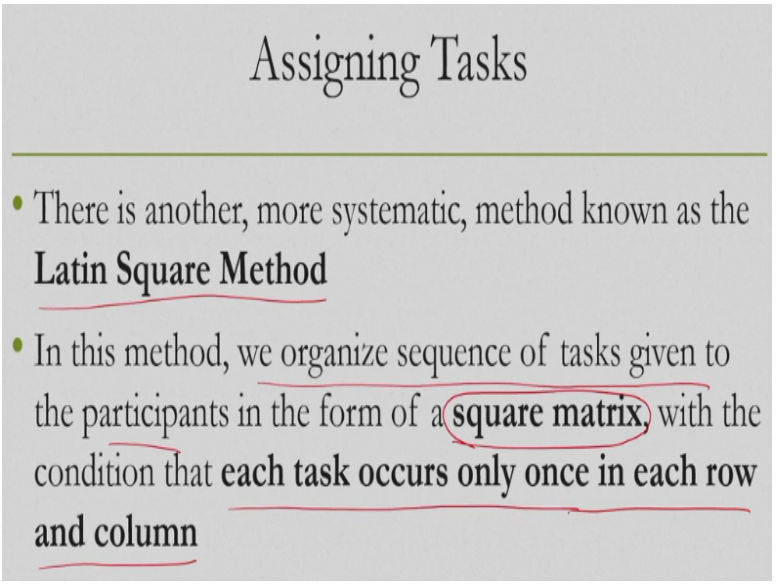
So, we do not provide the same sequence to each of the participants, for each we change the sequence so that practice effect is statistically minimized.

(Refer Slide Time: 41:00)



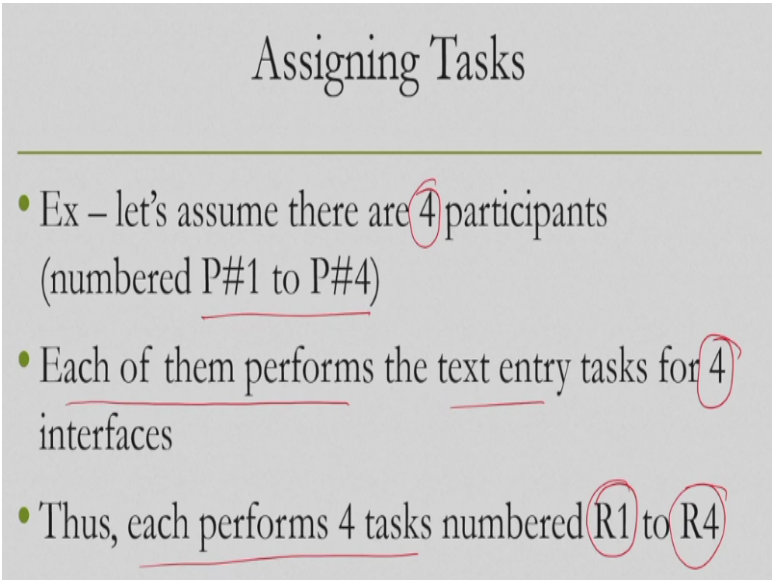
So, how we can implement the counterbalancing technique? We can use one popular technique that is randomizing the sequence for each participant. Now, of course randomization does not always lead to totally different sequence, it may lead to similar or partially similar sequence as well.

(Refer Slide Time: 41:23)



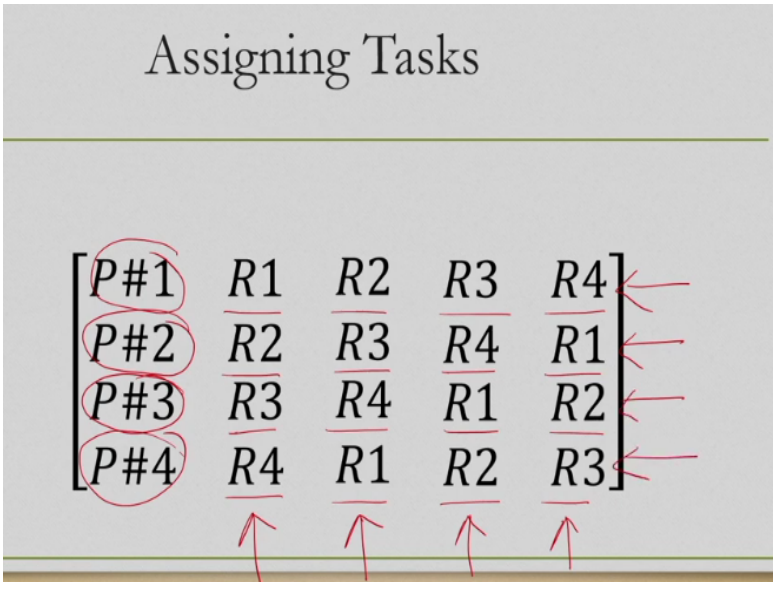
Another way to address this is to use a more systematic method known as the Latin Square method. In this method, what happens? We organize a sequence of tasks given to the participants in the form of a square matrix with the condition that each task occurs only once in each row and column. So, we form such a square matrix where each task occurs only once in each row and column. If we can frame such a matrix, then we can assign tasks following the matrix. Let us see one example.

(Refer Slide Time: 42:06)



Let us assume that there are 4 participants, participant number one 1 to participant number 4 and each of them performs the text entry tasks for 4 interfaces. So, there are 4 participants and 4 interfaces. So, each participant then performs 4 tasks numbered R1 to R4. Now, the question is how do we assign these tasks, in what sequence each participants get these tasks? If we follow the Latin Square method, how can we form the square and how the tasks are assigned?

(Refer Slide Time: 42:41)



This can be a possible assignment. So, P1 gets the task in the sequence R1 followed by R2 followed by R3 followed by R4. Then P2 gets the tasks in this sequence R2 followed by R3 followed by R4 followed by R1. P3 gets a different sequence R3 followed by R4 followed by R1 followed by R2 and P4 gets yet another sequence that is R4 followed by R1 followed by R2 followed by R3.

So, does it satisfy the condition that each row and column contain the task only once? Let us consider this row. So, here each task appears only once, so it satisfies the condition. And what about this column? So, here also we can see that each task appears only one, so it satisfies that condition. Similarly, this column also each task appears only once, in the third column also it holds, in the fourth column also it holds.

Similarly, second row it holds, third row it holds and fourth row also it holds. So, this is an example of a perfect Latin Square to assign tasks in a within-subject design to avoid practice effect and to implement counterbalancing. So, counterbalancing can be done in two ways, either we can randomize which may result in similar task sequences or to follow a more systematic method such as the Latin Square method.

So, these are the things that we should consider in the third stage of the empirical study that is how to design experiments. So, just to recap what we have learned here experiment design is not easy. It involves considerations for lots of issues. First issue is what kind of participants we should employ for our study. The idea is that the participants' profiles should match to the intended users' profile.

If the match is not there, then whatever conclusions we draw are applicable only to those users whose profile matches with the participants' profile. Second issue that we need to deal with is the number of participants what should be a good number to collect data. As we have seen 5 is a number suggested for carrying out pilot studies, but for a proper study we require between 12 to 25 participants, the more the better.

So, a number around 25 is probably a good number. Third issue that we need to consider is what kind of assignments it should follow. So, there are two ways we can design the experiments, one is within-subject, one is between-subject. In within-subject design also known as repeated-measure design, what we can do is we can ask each participant to participate in all the test conditions.

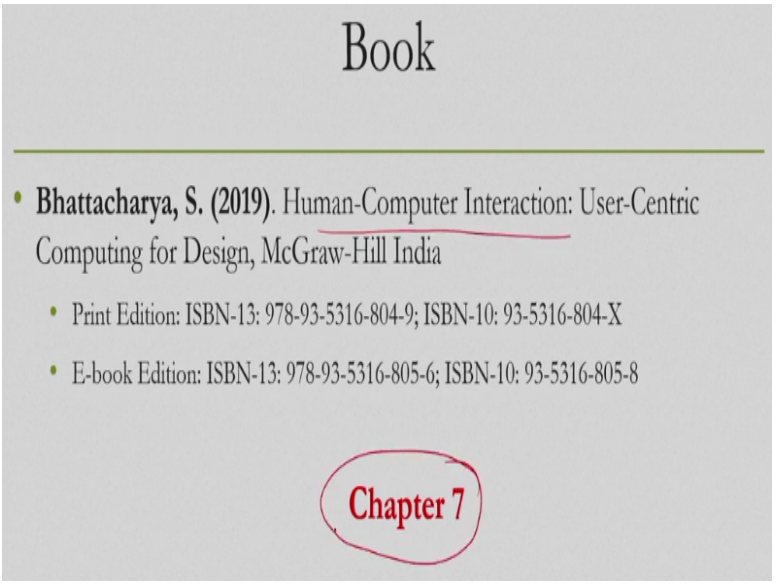
Now, if the number of test conditions are large or the participants have some problem, physical or professional, then probably that may not work., so we have to go for between-subject design. In between-subject design, we do not ask each participant to

participate in all the test conditions rather a subset of test conditions. Each design has its own pros and cons. If we go for within-subject design, then there is an issue of practice effect.

So, to avoid that we can go for either randomization of the task sequence in which the tasks are assigned to the participants or we can use systematic methods such as a Latin Square method. We have also talked about how to design tasks because tasks are important considerations in empirical study. So, we should be very careful in designing tasks, we should focus on representative tasks and design those tasks to be given to the participants.

I hope you have understood the concepts and enjoyed the content of this lecture. So, we will continue our discussion on the empirical study in the next lecture as well. So, far we have covered three stages, remaining stage is data analysis which we will take up in the next lecture. Looking forward to meet you all in the next lecture. Thank you and goodbye.

(Refer Slide Time: 47:25)



The material that we have covered today can be found in this book Human-Computer Interaction, chapter 7. That is all for this lecture. Thank you and goodbye.