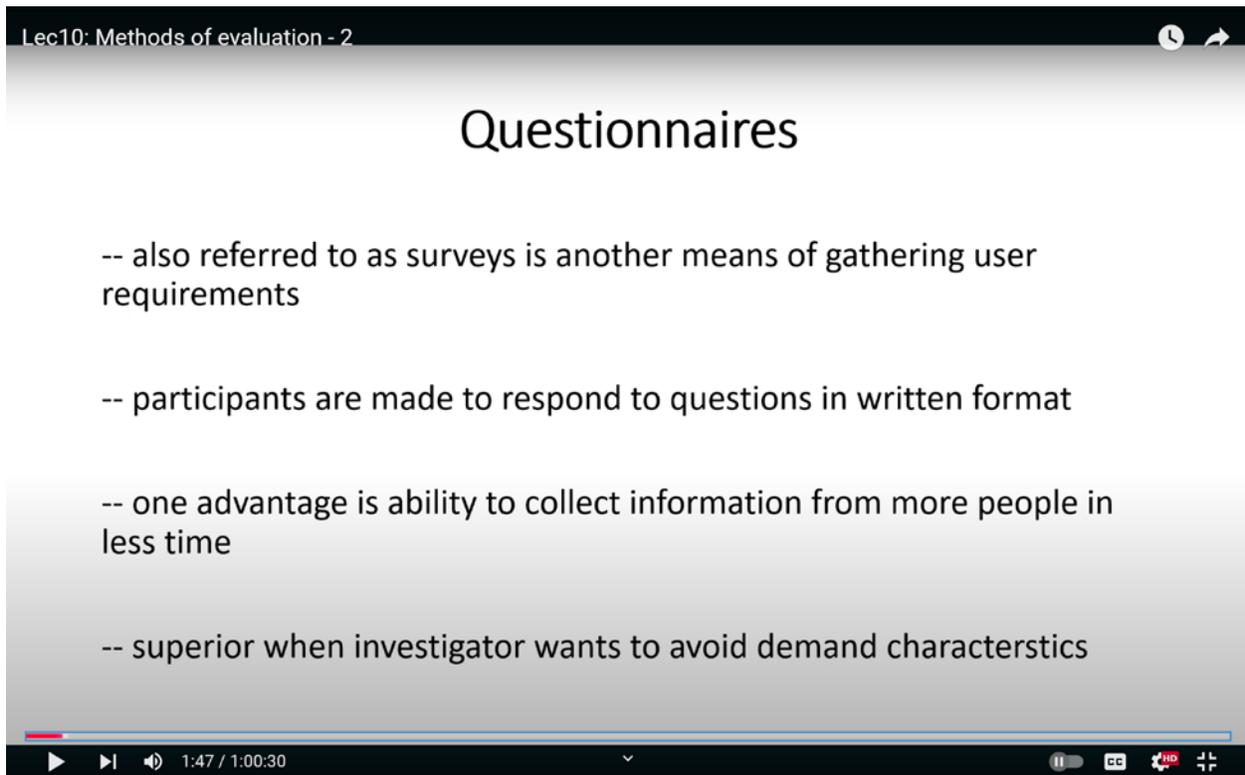


**Engineering Psychology**  
**Prof. Naveen Kashyap**  
**Department of Humanities and Social Sciences**  
**Indian Institute of Technology, Guwahati**  
**Week-04**  
**Lecture-10**  
**Methods of evaluation - 2**

Namaskar, and welcome to the second lecture on evaluation methods in engineering psychology. As you may recall from the first lecture, we discussed various evaluative methods, along with their advantages and disadvantages. We also explored user-centered design, emphasizing the importance of involving users at every stage of product development. This approach aids in creating better products with higher performance rates and lower error rates.

(Refer Slide Time: 01:47)



The screenshot shows a video player interface. At the top, the video title is "Lec10: Methods of evaluation - 2". The main content of the slide is titled "Questionnaires" and lists four bullet points: "-- also referred to as surveys is another means of gathering user requirements", "-- participants are made to respond to questions in written format", "-- one advantage is ability to collect information from more people in less time", and "-- superior when investigator wants to avoid demand characteristics". The video player controls at the bottom show a progress bar at 1:47 / 1:00:30, along with play, volume, and other standard controls.

Next, we examined the interview technique for collecting data from users participating in the user-centered design process. In today's lecture, we will delve deeper into additional methods for

gathering data related to user-centered design. We will also focus on task analysis and cognitive task analysis, and subsequently, we will explore methods for design evaluation.

Let us begin by addressing the nature of interviews. In interviews, there is a one-to-one interaction between an interviewer and a respondent, allowing for the collection of a substantial amount of data. While there are various forms of interviews, they can be quite expensive. Additionally, interviews often yield data that, although meaningful, can be highly varied, making it challenging to derive clear interpretations from the collected data. The sheer volume of data gathered in interviews can also pose a problem.

To address these challenges, questionnaires can be employed as an effective solution. Questionnaires are equivalent to surveys and serve as a structured format for collecting data from users. You may have encountered various questionnaires in your daily life. For instance, a website survey may contain questions about the performance of the designer's website. Alternatively, you might receive a questionnaire in a supermarket concerning the layout of the store or the availability of certain products. There may also be questionnaires distributed during flights that ask specific questions regarding the flight attendants, the atmosphere on the plane, and your overall experience and satisfaction with the flight. All questionnaires consist of a list of questions to which respondents are required to provide answers. Thus, questionnaires are a valuable tool for gathering user requirements and data.

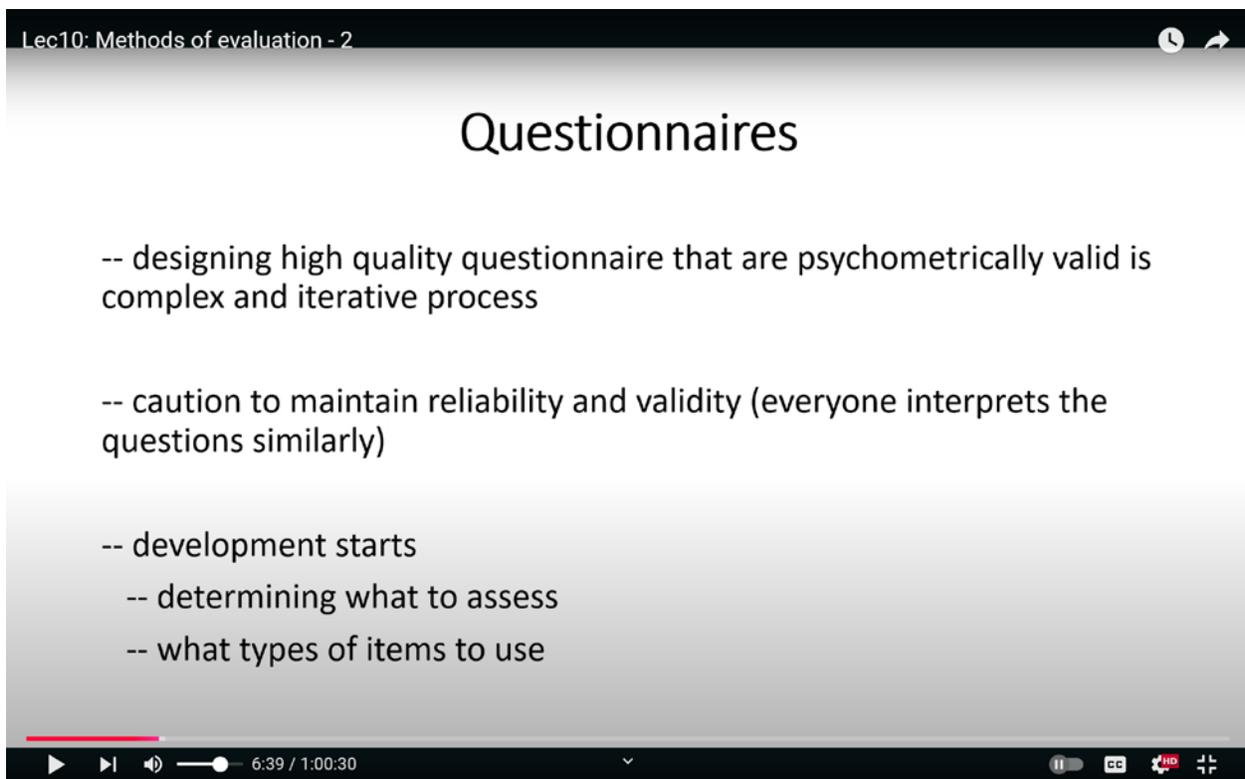
As previously mentioned, most questionnaires are in written form. Occasionally, you may come across auditory questionnaires, but the nature of auditory information is often transitory, making it difficult for respondents to retain substantial amounts of information. Consequently, verbal questionnaires have not proven to be very effective; written formats are far more common.

Interviews may struggle to collect large volumes of data, while questionnaires excel in this area. They can be distributed to many individuals in a short amount of time, allowing for the collection of a significant amount of data. When the quality of the data is essential, utilizing a questionnaire is advantageous.

Another important feature of questionnaires is that they help avoid demand characteristics. Demand characteristics refer to environmental factors that provide respondents with subtle cues

about how they should respond. For example, if someone is taking responses from you and unintentionally nods or smiles, their facial expressions may influence your answers. This type of environmental hint can guide users to provide specific types of responses, which is primarily a concern during one-on-one interviews. However, when filling out a questionnaire, the respondent does not have the interviewer present, thereby eliminating the issue of demand characteristics. Therefore, if you believe that experimental bias or demand characteristics may compromise your research, using a questionnaire would be a prudent choice.

(Refer Slide Time: 06:39)



Lec10: Methods of evaluation - 2

## Questionnaires

- designing high quality questionnaire that are psychometrically valid is complex and iterative process
- caution to maintain reliability and validity (everyone interprets the questions similarly)
- development starts
  - determining what to assess
  - what types of items to use

6:39 / 1:00:30

That said, designing questionnaires is not a straightforward task. A primary challenge is that questionnaires often represent the perspective of the question creator to the respondent. Thus, careful consideration must go into crafting a questionnaire. Questionnaires are complex instruments that require psychometric validation, and their development involves multiple iterations. When designing questions for a questionnaire, it is essential to ensure high validity and reliability, which should guide the selection of questions included within the questionnaire.

There are several features to consider when creating a questionnaire, which we will explore throughout this lecture. The necessity of maintaining psychometric validity in a questionnaire stems from the fact that different individuals may interpret the same question in diverse ways. For instance, if there is a question that seeks to define what life satisfaction means, older individuals might interpret life satisfaction as a sense of peace, reflecting on their accomplishments and fulfilling lives. In contrast, for someone aged 20 to 30, life satisfaction might be associated with acquiring more money to purchase new things, as these possessions and financial resources provide them with a sense of fulfillment. Consequently, life satisfaction can vary significantly among different demographic groups.

Therefore, these considerations should be taken into account when constructing a questionnaire. This is why it is crucial to maintain high levels of reliability and validity during the design phase of the questions. The development of a questionnaire typically begins with assessing and determining the specific aspect to be studied. This could involve examining performance, satisfaction, or another psychological attribute. Once the construct to be studied is clearly defined, one can proceed to create the questionnaire.

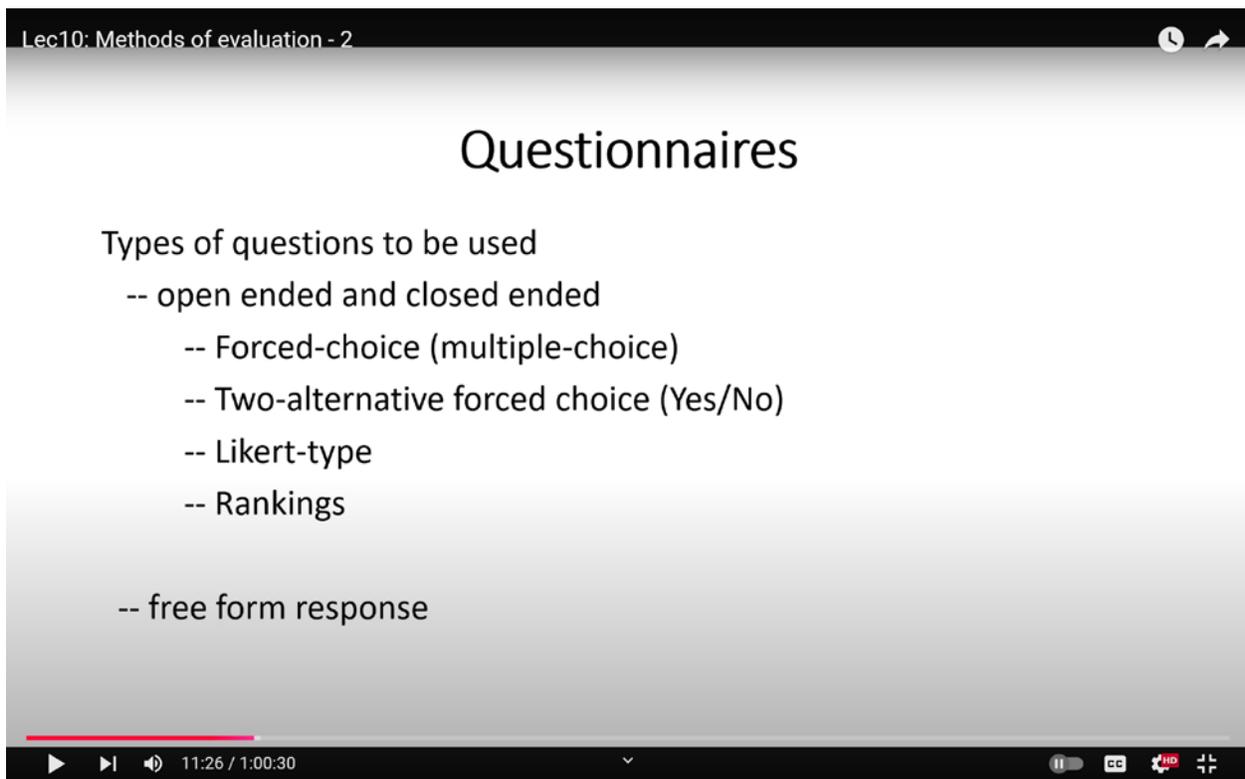
As mentioned, the first step in creating a questionnaire is to understand the construct you intend to study. The next step involves determining the types of measures that will be used. For example, if the goal is to understand happiness through a questionnaire, it is essential to recognize that happiness is a construct that may differ among individuals. When designing a happiness questionnaire, the creator must identify the various aspects of happiness that should be included. This could involve determining whether happiness is linked to specific personality traits, behavioral aspects, transcendental experiences, or other dimensions. Therefore, it is necessary to establish certain parameters that define happiness.

To gather these parameters, one should consult existing research. Researchers typically provide operationalizations when defining constructs like happiness. During this process, they present their interpretations and definitions of happiness. By spending sufficient time on this research, one can identify the factors that define happiness and incorporate these factors into the questionnaire to measure it effectively. This process marks the beginning of questionnaire development.

Questionnaires can feature a variety of question types. Two significant categories of questions that

may be included are open-ended and closed-ended questions. Open-ended questions might ask respondents to define how satisfied they feel after traveling with a particular airline. Such questions allow respondents to provide any answer related to their experiences. Conversely, if a quick response is desired regarding fixed values, a closed-ended question might be posed, such as asking respondents to rate their satisfaction while traveling on a scale of 1 to 5. Alternatively, multiple-choice questions or yes/no questions can be employed, requiring respondents to select an answer.

(Refer Slide Time: 11:26)



Lec10: Methods of evaluation - 2

## Questionnaires

Types of questions to be used

- open ended and closed ended
  - Forced-choice (multiple-choice)
  - Two-alternative forced choice (Yes/No)
  - Likert-type
  - Rankings
- free form response

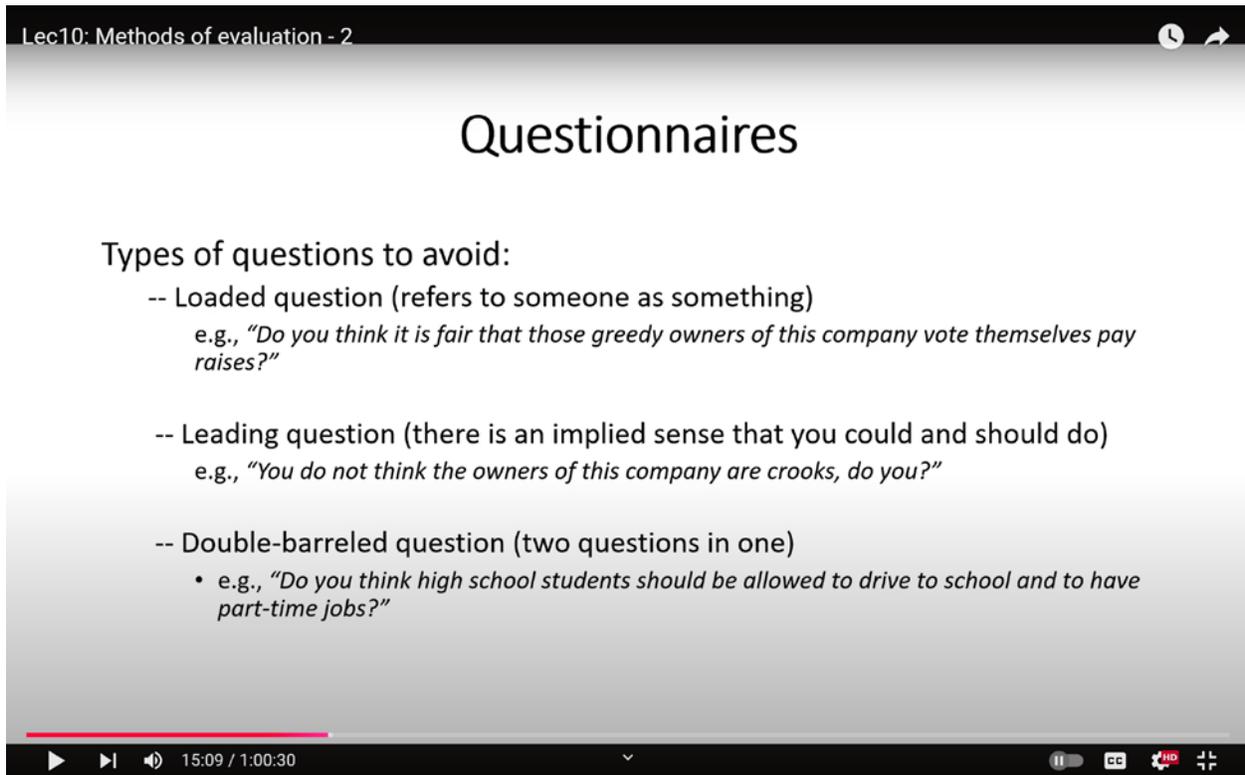
11:26 / 1:00:30

Most questionnaires can incorporate both open-ended and closed-ended questions. Closed-ended questions can take the form of forced-choice response types, such as multiple-choice questions where four options are provided. In this scenario, one or more answers may be correct, while the others serve as distractors. The task for the respondent is to identify the correct option. Recognition is the type of memory employed in answering such questions.

Two-alternative forced-choice questions can also be utilized, where user satisfaction is rated in binary terms (yes or no). Another common format is the Likert-type scale, which benchmarks

satisfaction by asking respondents to evaluate their experiences traveling on a scale of 1 to 5, where 0 indicates no satisfaction and 5 represents extreme satisfaction. This type of response requires individuals to assess their level of satisfaction and select a corresponding value.

(Refer Slide Time: 15:09)



Lec10: Methods of evaluation - 2

## Questionnaires

Types of questions to avoid:

- Loaded question (refers to someone as something)  
e.g., *"Do you think it is fair that those greedy owners of this company vote themselves pay raises?"*
- Leading question (there is an implied sense that you could and should do)  
e.g., *"You do not think the owners of this company are crooks, do you?"*
- Double-barreled question (two questions in one)
  - e.g., *"Do you think high school students should be allowed to drive to school and to have part-time jobs?"*

15:09 / 1:00:30

Ranking scales allow respondents to assign ranks to their satisfaction levels during travel, enabling them to prioritize among various alternatives. Additionally, free-form responses can be included in a questionnaire, resembling open-ended questions but providing some guidance on what should be addressed in these responses. Participants are free to articulate their thoughts in any manner they prefer.

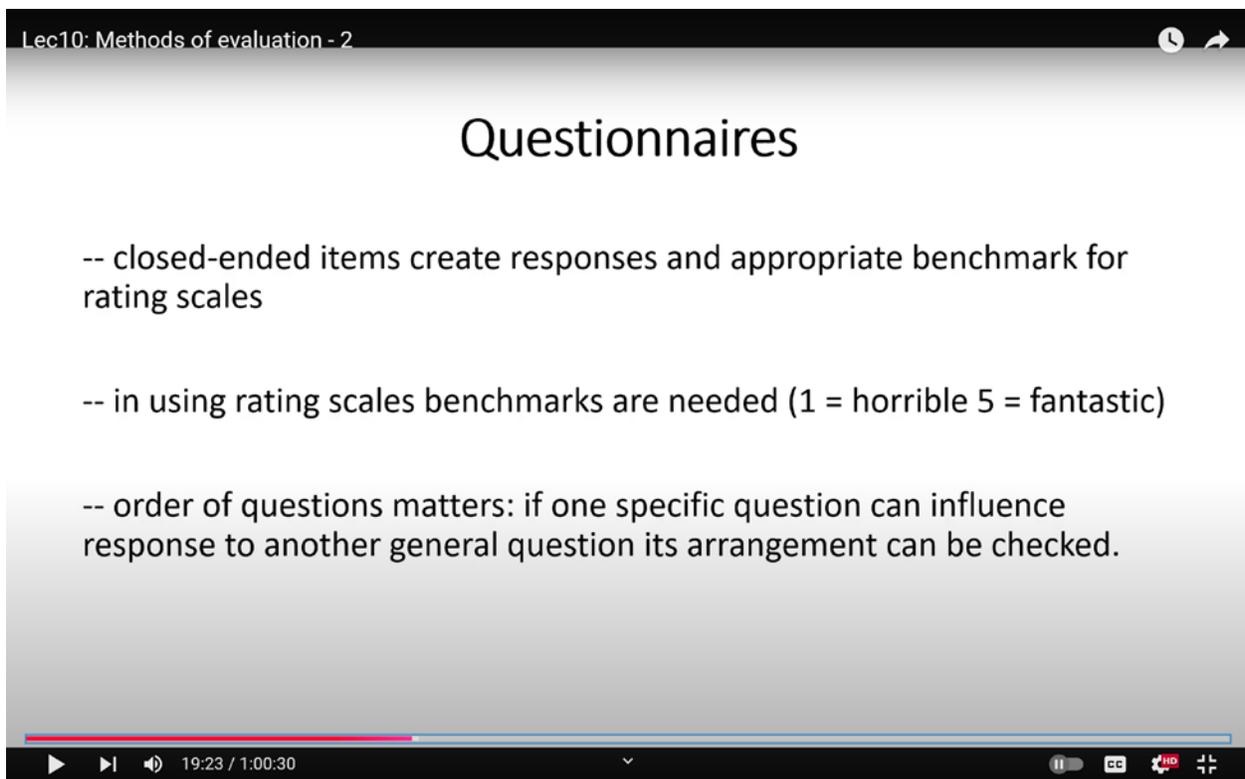
While designing a questionnaire, it is essential to be cautious about three key aspects. The first aspect to consider is loaded questions.

When creating a questionnaire, it is crucial to avoid formulating loaded questions. A loaded question is one that provides a hint or an implication that influences how respondents perceive and answer it. For example, consider the question: "Do you think it is fair for these greedy owners of

this company to vote for their own pay raises?" The term "greedy owners" inherently suggests that the owners are selfish, which introduces a bias that may affect how respondents think and answer the question.

Additionally, questionnaires should not contain leading questions. A leading question implies a particular response or direction that respondents should take. For instance, the question "You don't think the owners of this company are crooks, do you?" implies that the owners might indeed be crooks, which can sway respondents to think along those lines when formulating their answers.

(Refer Slide Time: 19:23)



Lec10: Methods of evaluation - 2

## Questionnaires

- closed-ended items create responses and appropriate benchmark for rating scales
- in using rating scales benchmarks are needed (1 = horrible 5 = fantastic)
- order of questions matters: if one specific question can influence response to another general question its arrangement can be checked.

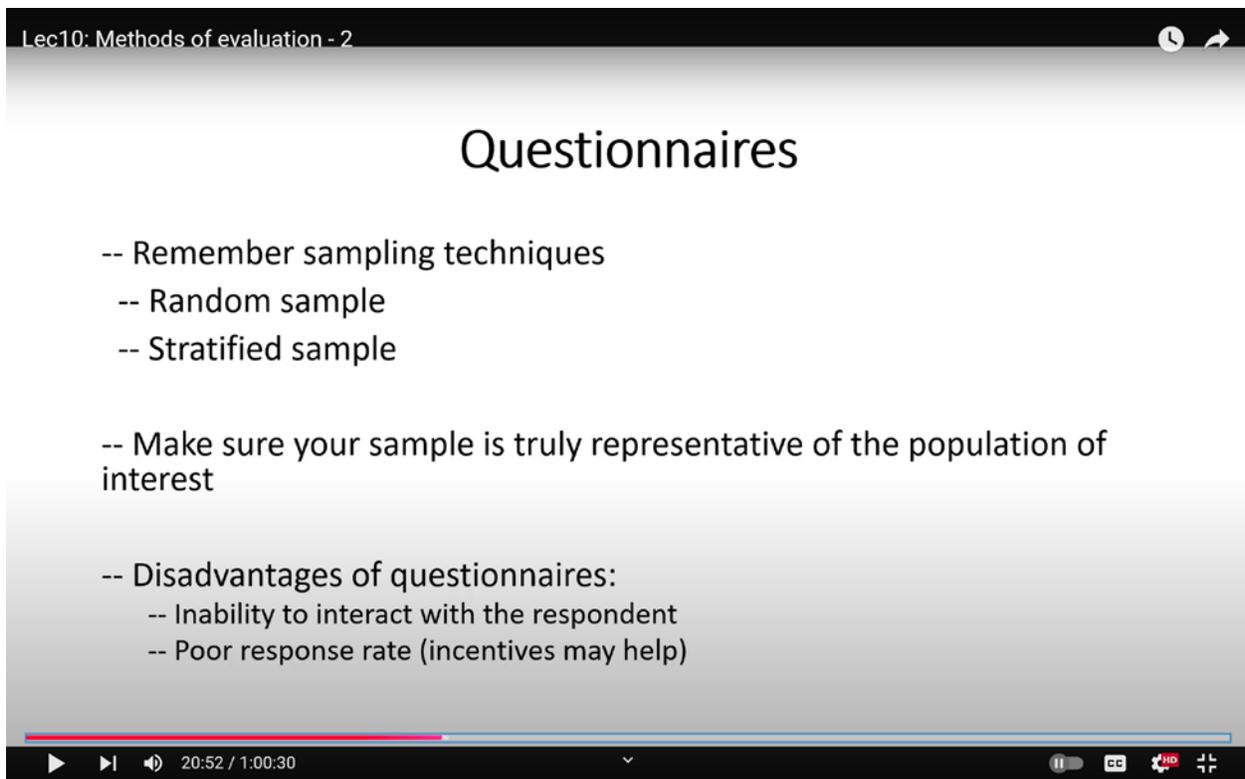
19:23 / 1:00:30

Another type of question to avoid is the double-barreled question, which combines two inquiries into a single question, making it unclear which part the respondent should prioritize in their answer. For example, the question "Do you think high school students should be allowed to drive to school and have part-time jobs?" confuses respondents regarding whether they should answer about driving, part-time jobs, or both simultaneously, complicating their ability to provide a coherent response.

Closed-ended items in a questionnaire require fixed responses, and it is essential to establish a benchmark for these responses. Without a benchmark or an internal scale for rating answers, it can be challenging for respondents to provide meaningful feedback. For example, if you do not specify the scale on which satisfaction will be rated, whether it is a one-directional scale that ranges from "good" to "excellent," or a two-directional scale that extends from "not at all" to "extreme," with "neutral" in the center, it can lead to confusion.

Providing clear benchmarks for rating scales is vital, as it minimizes the need for respondents to think too deeply about their answers. An example of a rating scale could range from 1, denoting "horrible," to 5, denoting "fantastic," thereby guiding respondents from extremely negative to extremely positive evaluations.

(Refer Slide Time: 20:52)



Lec10: Methods of evaluation - 2

## Questionnaires

- Remember sampling techniques
  - Random sample
  - Stratified sample
- Make sure your sample is truly representative of the population of interest
- Disadvantages of questionnaires:
  - Inability to interact with the respondent
  - Poor response rate (incentives may help)

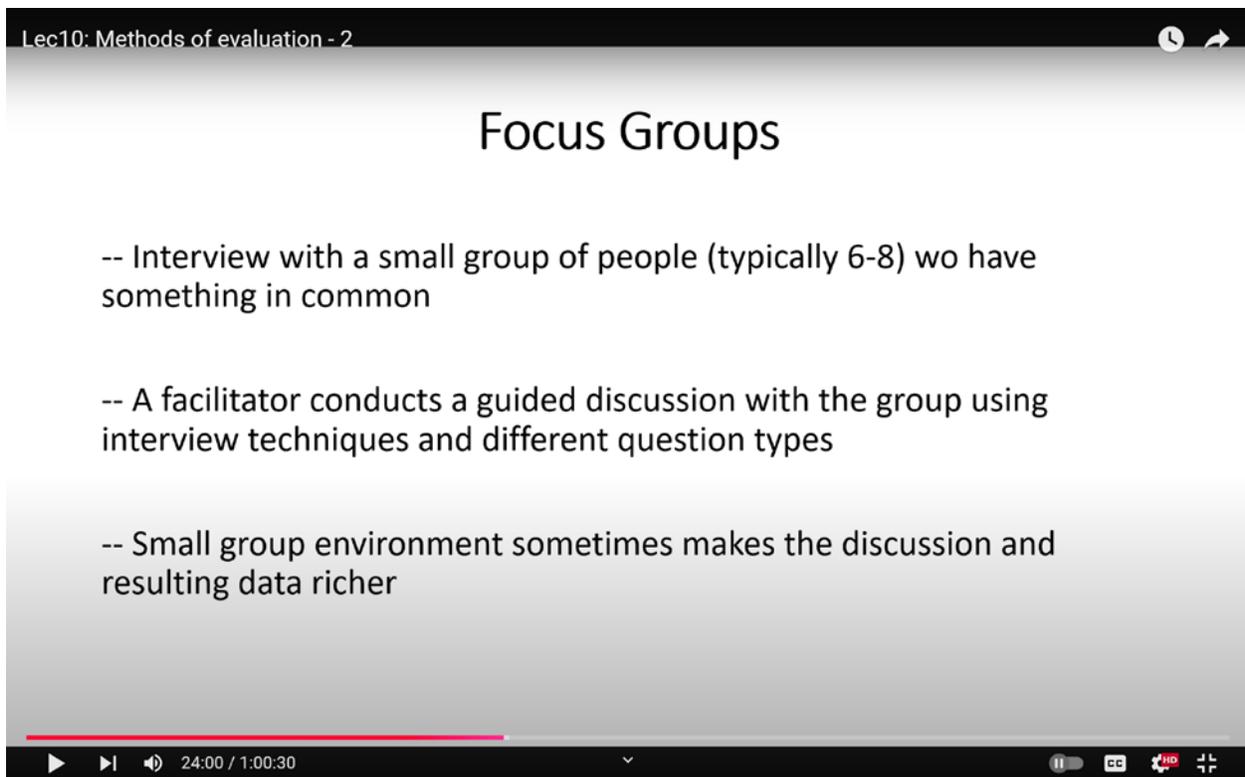
20:52 / 1:00:30

The order of questions is another important consideration in questionnaire design. Sometimes, the way questions are structured can influence how respondents answer subsequent questions. If one specific question inadvertently provides insight into another more general question, this can skew

the responses. This issue is particularly pertinent in paper-and-pencil tests, where respondents might go back to adjust their answers after considering the influence of earlier questions. Therefore, careful arrangement of questions is essential.

Recalling the sampling techniques discussed in the earlier sections of this course, such as random and stratified sampling, is important for questionnaire design. Whether the questionnaire is targeted at a specific group or the general population will influence how the questions should be framed. A key principle in questionnaire design is ensuring that the sample collected represents the population. For instance, if a questionnaire is designed to assess the effectiveness of a product, it is not feasible to study everyone who has used the product. Instead, a representative sample must be selected, encompassing various demographics of the product's users.

(Refer Slide Time: 24:00)



Lec10: Methods of evaluation - 2

## Focus Groups

- Interview with a small group of people (typically 6-8) who have something in common
- A facilitator conducts a guided discussion with the group using interview techniques and different question types
- Small group environment sometimes makes the discussion and resulting data richer

24:00 / 1:00:30

Despite their advantages, questionnaires have notable disadvantages. One significant drawback is the lack of interaction with respondents. Since the researcher cannot be present when respondents answer the questions, there is little opportunity to clarify or elaborate on the questions. This

limitation increases the risk of misinterpretation, which can skew the results. Additionally, questionnaires often suffer from poor response rates, further complicating data collection efforts.

Most individuals do not take questionnaires very seriously, which often results in a reluctance to submit them. Consequently, researchers frequently find that the number of respondents returning completed questionnaires is quite low. This decline in response rates presents a significant challenge when utilizing questionnaires for data collection.

An alternative method for collecting data is through focus groups. In a focus group, a small group of individuals, typically comprising 6 to 8 people, who share certain common characteristics, are gathered in a room. This group serves as a representative sample of the target population. Discussions are held with these participants to gather as much information as possible, particularly from those who will be using the modified design or new software being evaluated.

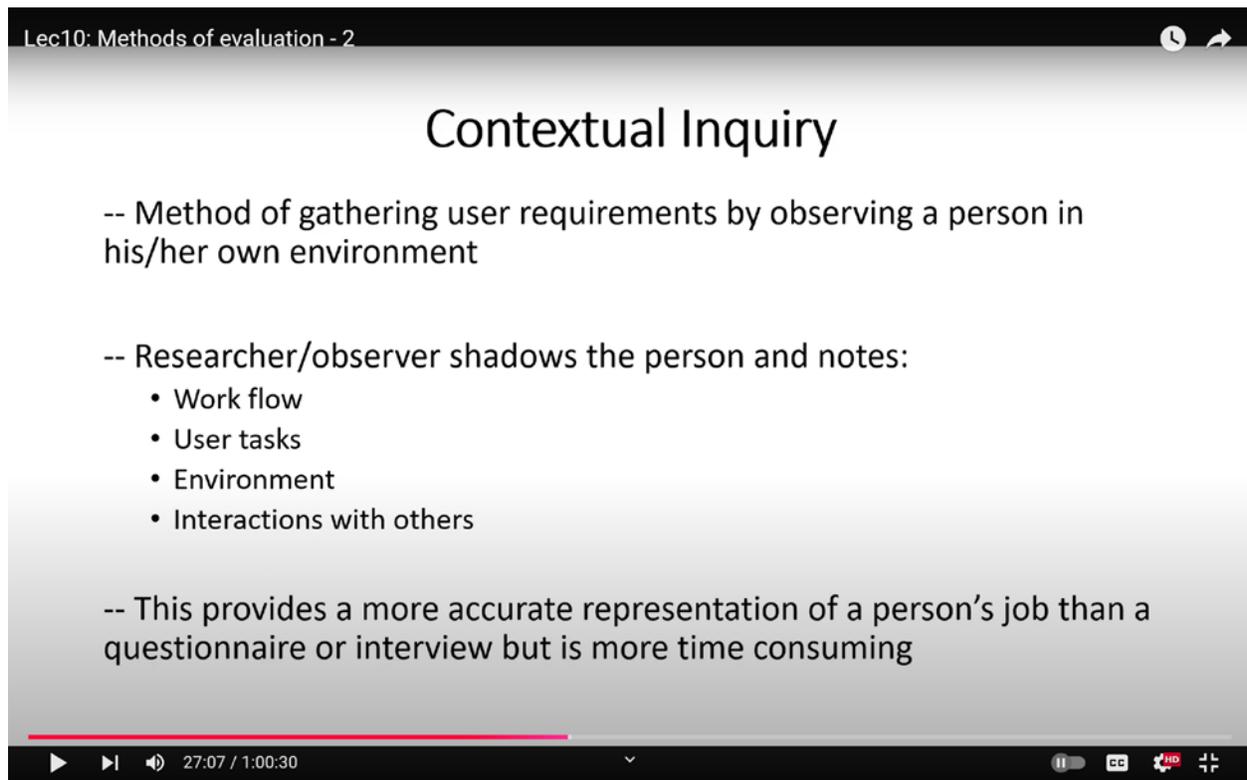
As previously discussed in the context of interviews, a focus group can involve various configurations: multiple respondents with a single interviewer, a single respondent with multiple interviewers, or a combination of both. This setup allows for a dynamic interaction among participants and the facilitator.

Focus groups are particularly effective in providing specific answers, and there are several approaches to creating one. Typically, a facilitator conducts a guided discussion using various interview techniques and question types. In this context, the facilitator leads the group in a one-on-one discussion, encouraging participants to share their thoughts and experiences.

The data obtained from focus groups tends to be of higher quality than that from traditional interviews because the participants are users of the software or product being discussed. Engaging with a small, diverse group increases the likelihood of receiving constructive and informative feedback.

Another technique for data collection that is similar to focus groups, but with notable differences, is called contextual inquiry. In this approach, researchers observe individuals in their natural work environments and gather data based on these observations. Researchers engage in discussions with the participants, combining observational data with qualitative insights from the discussions to create a comprehensive understanding of user needs.

(Refer Slide Time: 27:07)



Lec10: Methods of evaluation - 2

## Contextual Inquiry

- Method of gathering user requirements by observing a person in his/her own environment
- Researcher/observer shadows the person and notes:
  - Work flow
  - User tasks
  - Environment
  - Interactions with others
- This provides a more accurate representation of a person's job than a questionnaire or interview but is more time consuming

27:07 / 1:00:30

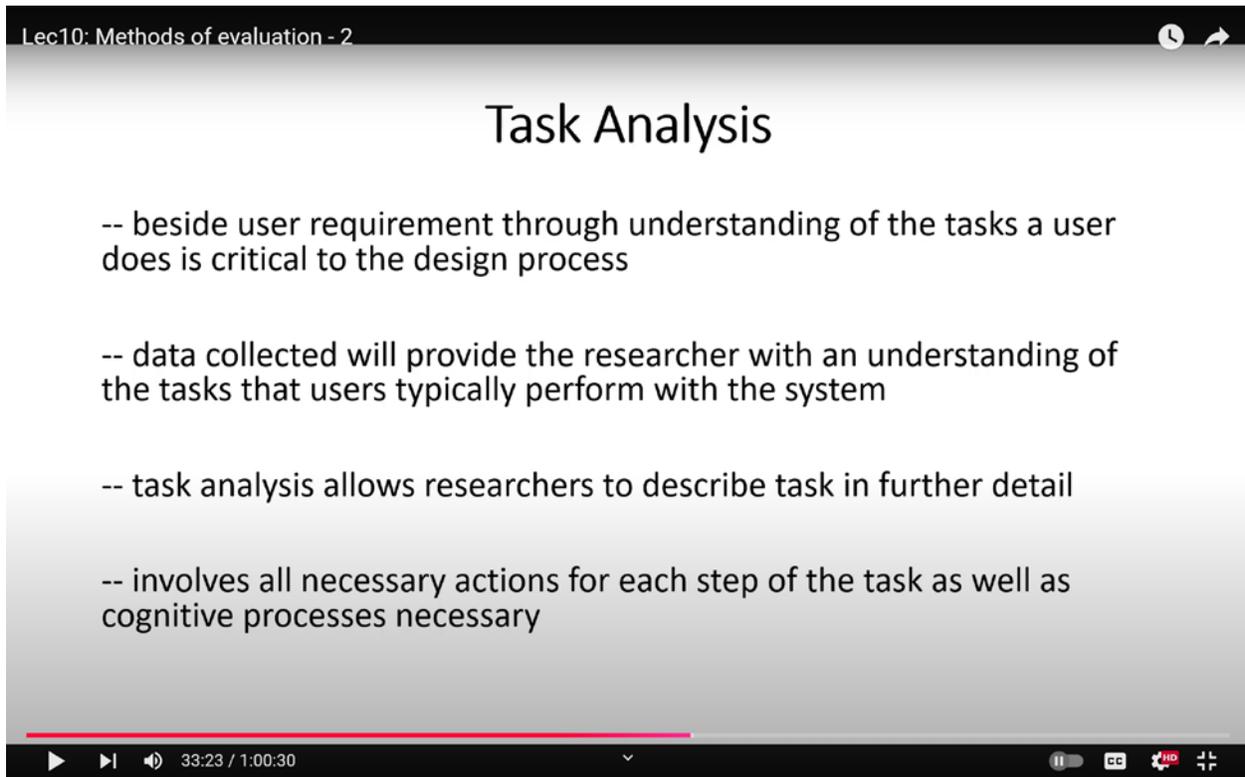
One significant advantage of contextual inquiry is that it allows researchers to capture authentic, natural responses from individuals as they perform their tasks. Unlike the artificial setting of a focus group, participants in a contextual inquiry behave as they normally would in their environments, revealing errors and shortcuts in their interactions with products or tasks.

During a contextual inquiry, researchers observe and shadow participants, taking detailed notes about their workflows, the steps involved in completing their tasks, and the specific actions they take in various assignments. For instance, when composing an email, a user must log in, open the email client, locate the compose button, enter a recipient's address, type a subject line, draft the message body, review it, and then send it. Researchers document all these steps while observing participants in context.

The researchers also take note of the surrounding environment, as it can significantly impact performance. Factors such as high temperatures, inadequate lighting, and excessive noise can hinder a user's ability to complete tasks effectively. Understanding the environment where users

work provides crucial insights for conducting task analyses and measuring the impact of design changes.

(Refer Slide Time: 33:23)



Lec10: Methods of evaluation - 2

## Task Analysis

- beside user requirement through understanding of the tasks a user does is critical to the design process
- data collected will provide the researcher with an understanding of the tasks that users typically perform with the system
- task analysis allows researchers to describe task in further detail
- involves all necessary actions for each step of the task as well as cognitive processes necessary

33:23 / 1:00:30

Finally, researchers collect data on how users interact with one another and with senior management, as well as their engagement with the system. This information offers valuable clues about the natural workflow of tasks and highlights potential error points that need to be addressed.

While contextual inquiry provides a more accurate representation of an individual's job than questionnaires or interviews, it is more time-consuming. The process involves spending extended periods with participants, conducting observations, and engaging in discussions, resulting in high-quality, rich data that can significantly inform design and usability improvements.

In an interview or a questionnaire, the process is typically a one-time event. An interview takes some time to conduct, but once it is finished, the interaction concludes. Conversely, a questionnaire is much simpler; it consists of questions to which the respondent provides answers.

On one hand, contextual inquiry is a data-rich form of investigation; however, it loses its advantage in terms of time and cost-effectiveness. We have discussed various methods for data collection, and the next step is to conduct a task analysis. The central idea behind this section is to evaluate designs and assess modifications or changes, whether they are new or proposed for existing systems. Task analysis serves as a vital tool in this evaluation process.

Task analysis provides developers or research teams in a manufacturing setting with insights into what the tasks are, how they are performed, and the specific points at which errors may occur. By conducting a thorough task analysis, followed by cognitive task analysis, developers and research teams gain an understanding of the natural flow of a task. They learn about the mental models that users employ while completing tasks and the shortcuts they utilize. This combination of information aids developers in designing modifications that could enhance the efficiency of older designs or new products.

Therefore, comprehending user requirements and understanding the tasks that users perform is critical to the design process. The data collected will give researchers insights into the tasks users typically perform with the system. Information gathered from interviews, contextual inquiries, or focus groups will reveal what users commonly do and how they complete their tasks.

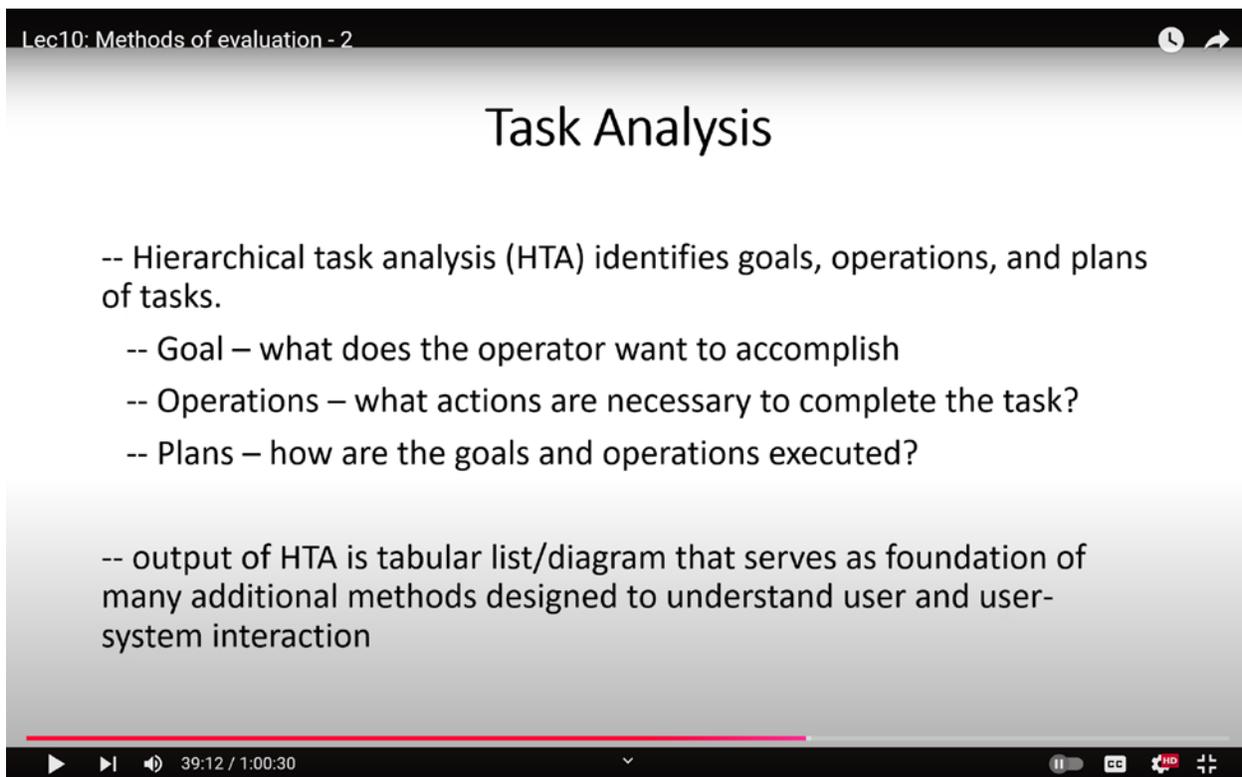
For instance, consider an operator working at a power generation plant. Through interviews, focus groups, or contextual inquiries, the researcher can observe how the operator performs their tasks and how they can demonstrate this to the researcher. By examining the operator's workflow in a power plant, researchers can identify areas that require modification in terms of systems or design interfaces. They will gain insight into the natural sequence of job execution within a power plant.

By collecting data through discussions and observational notes, researchers can identify specific points that hinder task execution and reduce performance. This information allows them to address error points in future design modifications. Task analysis specifically enables researchers to describe tasks in greater detail. While observing a worker in a power plant may reveal their mental models, engaging in dialogue with these workers uncovers additional information that cannot be gathered through observation alone. This is a key benefit of task analysis, as it highlights discrepancies between how tasks are performed and how they should ideally be performed.

Task analysis involves detailing all necessary actions for each step of a task, along with the cognitive processes required. For example, in a power plant, there are specific steps that must be followed, such as initiating the inflow of energy needed to run turbines. These turbines generate electricity, which is then transmitted to the grid. All of these steps must be executed in a sequential manner.

Through task analysis, researchers learn not only how operators perform their jobs but also what they are thinking during specific tasks. It is often observed that operators create mental shortcuts to complete tasks more quickly and with less effort, despite established procedures. Identifying these shortcuts and understanding the factors that contribute to their development can provide valuable insights for designing improved user interfaces in future systems. This information can be gathered through cognitive task analysis or cognitive process analysis.

(Refer Slide Time: 39:12)



Lec10: Methods of evaluation - 2

## Task Analysis

- Hierarchical task analysis (HTA) identifies goals, operations, and plans of tasks.
  - Goal – what does the operator want to accomplish
  - Operations – what actions are necessary to complete the task?
  - Plans – how are the goals and operations executed?
- output of HTA is tabular list/diagram that serves as foundation of many additional methods designed to understand user and user-system interaction

39:12 / 1:00:30

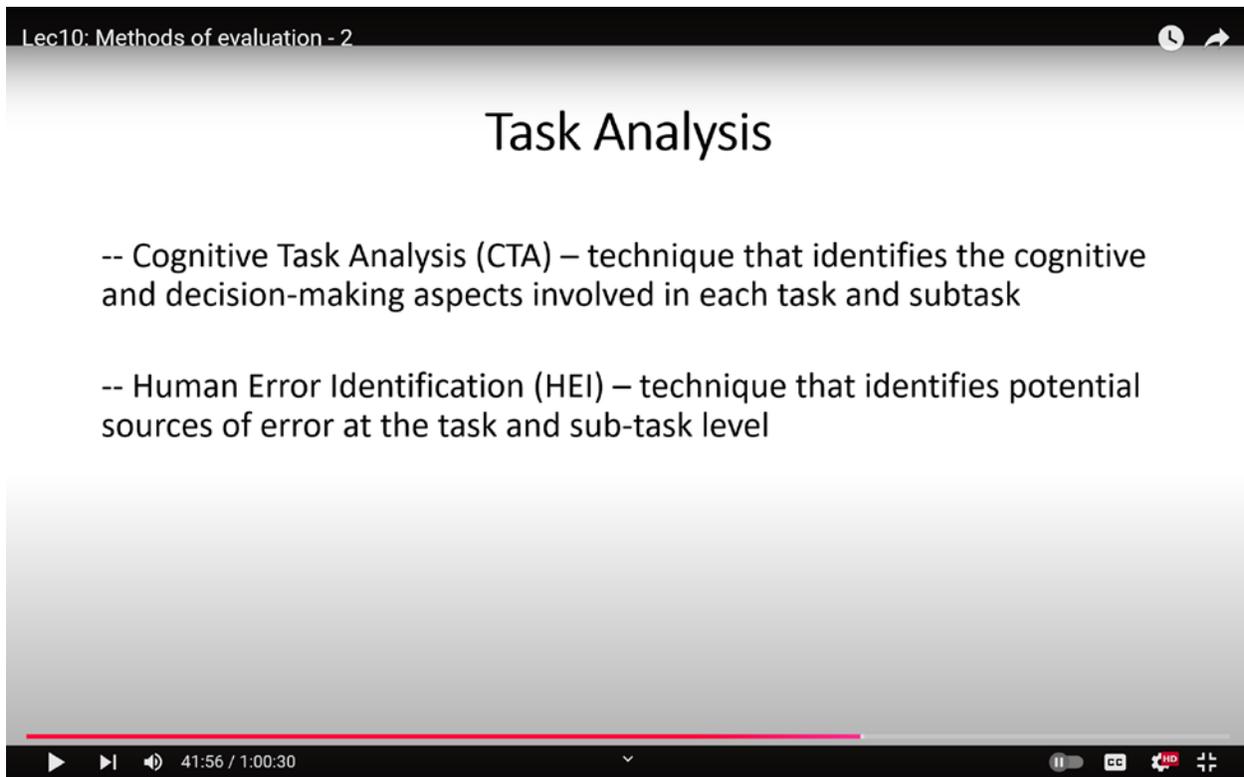
Task analysis is conducted using a method called hierarchical task analysis. This method comprises three components: the identification of goals, which outlines what is to be achieved; the

identification of operations, which describes what actions are necessary to achieve the goals; and the identification of plans, which details the step-by-step procedures required to execute the operations and attain the goals.

Hierarchical task analysis will clarify what the goals are, what the operator aims to accomplish, along with the operations required to complete the task and the plans for executing these operations. For instance, when writing an email, the goal is to send the email. The operations include logging in, finding the compose button, entering the recipient's address, typing the message, and sending it. The plan consists of the step-by-step actions needed to achieve the goal of sending the email.

How do you log in? Where do you search for the compose button? Is there a shortcut to it? If there is, what information should be provided in the compose box? The minimum information needed to send an email includes which fields are necessary, where the send button is, and all this information constitutes the plan.

(Refer Slide Time: 41:56)



Lec10: Methods of evaluation - 2

## Task Analysis

- Cognitive Task Analysis (CTA) – technique that identifies the cognitive and decision-making aspects involved in each task and subtask
- Human Error Identification (HEI) – technique that identifies potential sources of error at the task and sub-task level

41:56 / 1:00:30

This illustrates how the goals, operations, and plans of any job can be organized using hierarchical task analysis. The output of a hierarchical task analysis is presented in either a tabular form or a diagram, which serves as the foundation for various additional methods aimed at understanding user interactions with systems. The goals, operations, and plans are organized in a way that allows researchers or developers to refer to the table and identify areas that require improvement. They can also pinpoint components that contribute to errors in task completion or that may pose potential hazards during the execution of tasks. Once developers have this information, they can rethink the easiest way to write an email. This is the purpose of conducting hierarchical task analysis: to provide clear, organized information in tabular form regarding all the steps involved in completing a job.

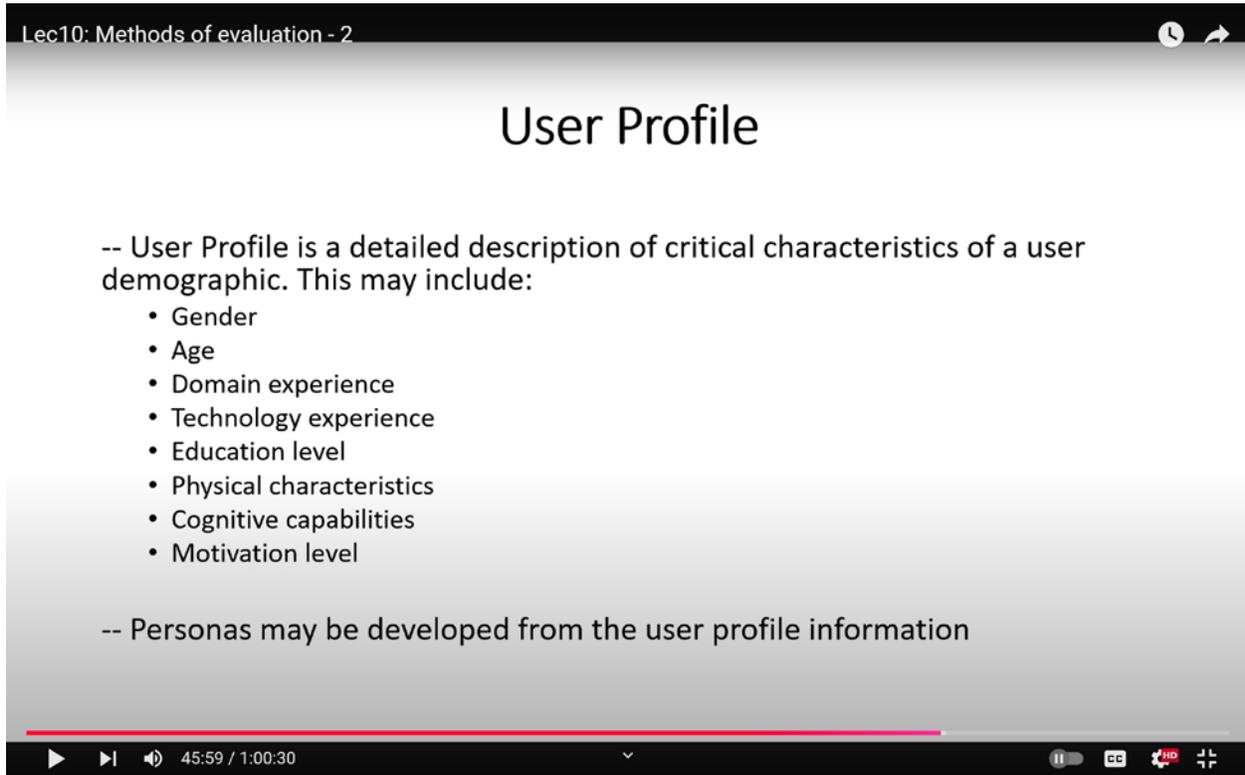
Cognitive task analysis is a technique that identifies the cognitive and decision-making aspects involved in each task and subtask. It delves into the thought processes behind writing an email. The concept of email is derived from traditional postal mail or letters, which users are familiar with. People have an understanding of what elements are necessary in postal mail: an address is essential, some text is important, and an envelope or cover is required, along with a method of sending it.

This mental model can be applied to the creation of emails as well. In an email, there must be a body, an address, and a method of sending it, which involves clicking "send." The decisions and thought processes that users have regarding writing an email can be utilized to enhance how people write emails or interact with an email program. The thoughts, plans, and decisions individuals have when starting to compose an email can be outlined, leading to improvements that result in better designs.

Another valuable insight derived from hierarchical task analysis is human error identification. Through task analysis, researchers can discover where users make errors, what factors contribute to these mistakes, and the specific points at which errors may occur. With this knowledge, they can develop better designs that incorporate more robust error-handling mechanisms. Techniques that identify potential sources of error at both the task and subtask levels can significantly enhance usability. For example, placing the compose button prominently at the top of the email interface, where users are likely to look first, can reduce errors.

Similarly, positioning the send button at the top facilitates easy access after users review their emails. Conversely, if the send button is hidden or placed in an unexpected location, it may lead to errors. These aspects can be studied to develop a modified and efficient email system.

(Refer Slide Time: 45:59)



The image is a screenshot of a video slide. At the top, there is a black header bar with the text 'Lec10: Methods of evaluation - 2' on the left and a clock icon on the right. Below the header, the title 'User Profile' is centered in a large, black, sans-serif font. The main content area has a light gray background and contains two bullet points. The first bullet point is a double dash followed by the text: 'User Profile is a detailed description of critical characteristics of a user demographic. This may include:'. Below this is a list of seven items, each preceded by a single bullet point: Gender, Age, Domain experience, Technology experience, Education level, Physical characteristics, Cognitive capabilities, and Motivation level. The second bullet point is a double dash followed by the text: 'Personas may be developed from the user profile information'. At the bottom of the slide, there is a video player interface with a red progress bar, a play button, a volume icon, and a timestamp '45:59 / 1:00:30'. On the right side of the player, there are icons for a closed caption, a red 'HD' logo, and a full screen icon.

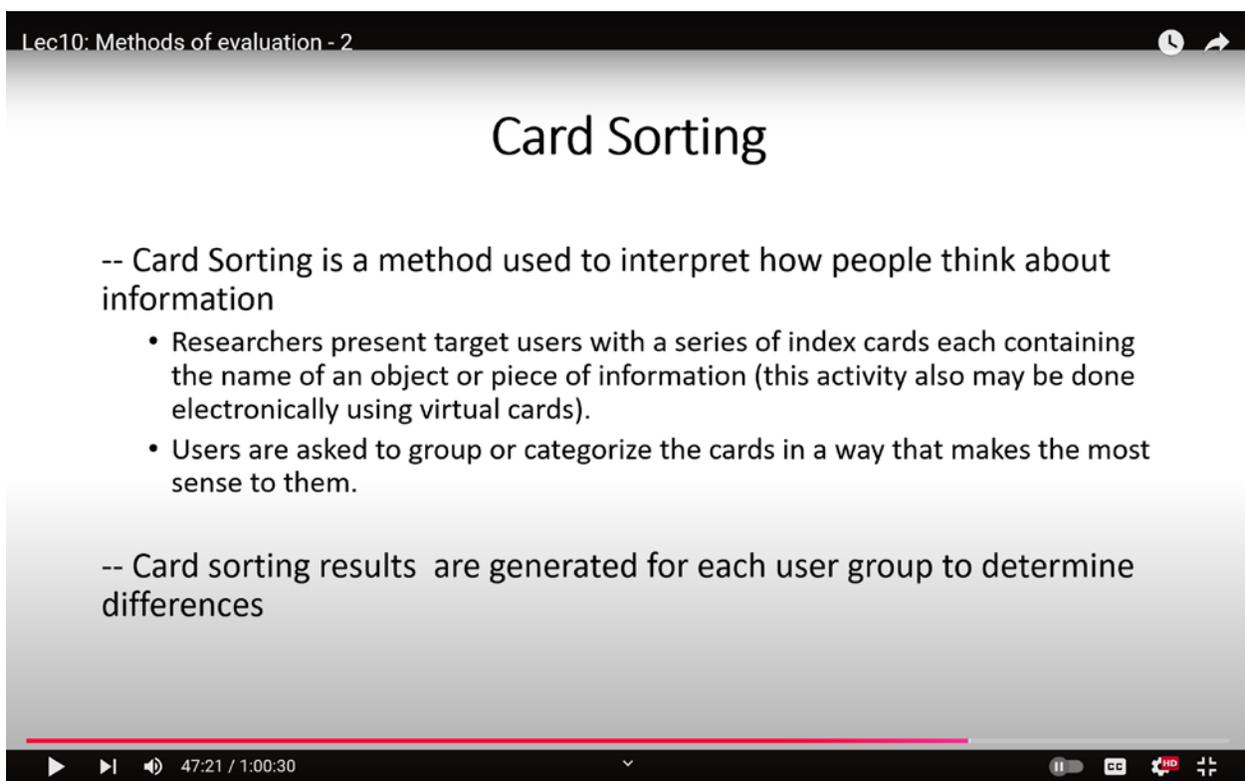
Following task analysis and prior to data collection, a user profile analysis is conducted. What does user profile analysis entail? When collecting data, it is essential to understand who the target group is. User profile analysis provides a detailed description of the critical characteristics of the user demographic. Researchers need to be aware of factors such as gender, age, domain experience, technological experience, education level, physical characteristics, cognitive capabilities, and motivation.

This knowledge allows researchers to better understand who their users are, which in turn aids in the design process. From user profile information, personas can be developed. Once the user demographics are established, artificial personas can be created. For example, if a new type of face cream is introduced, the researcher must determine the target audience for the product. Face creams

are particularly popular among certain age groups, which may vary in their purchasing motivations and interests.

By creating personas representing different demographics, such as age, education level, financial status, and motivation behind using the cream, the researcher can tailor the product to meet the needs of these hypothetical users. Personas are fictional representations created based on data gathered from the target user group.

(Refer Slide Time: 47:21)



The image is a screenshot of a video player showing a slide titled "Card Sorting". The slide content is as follows:

Lec10: Methods of evaluation - 2

## Card Sorting

-- Card Sorting is a method used to interpret how people think about information

- Researchers present target users with a series of index cards each containing the name of an object or piece of information (this activity also may be done electronically using virtual cards).
- Users are asked to group or categorize the cards in a way that makes the most sense to them.

-- Card sorting results are generated for each user group to determine differences

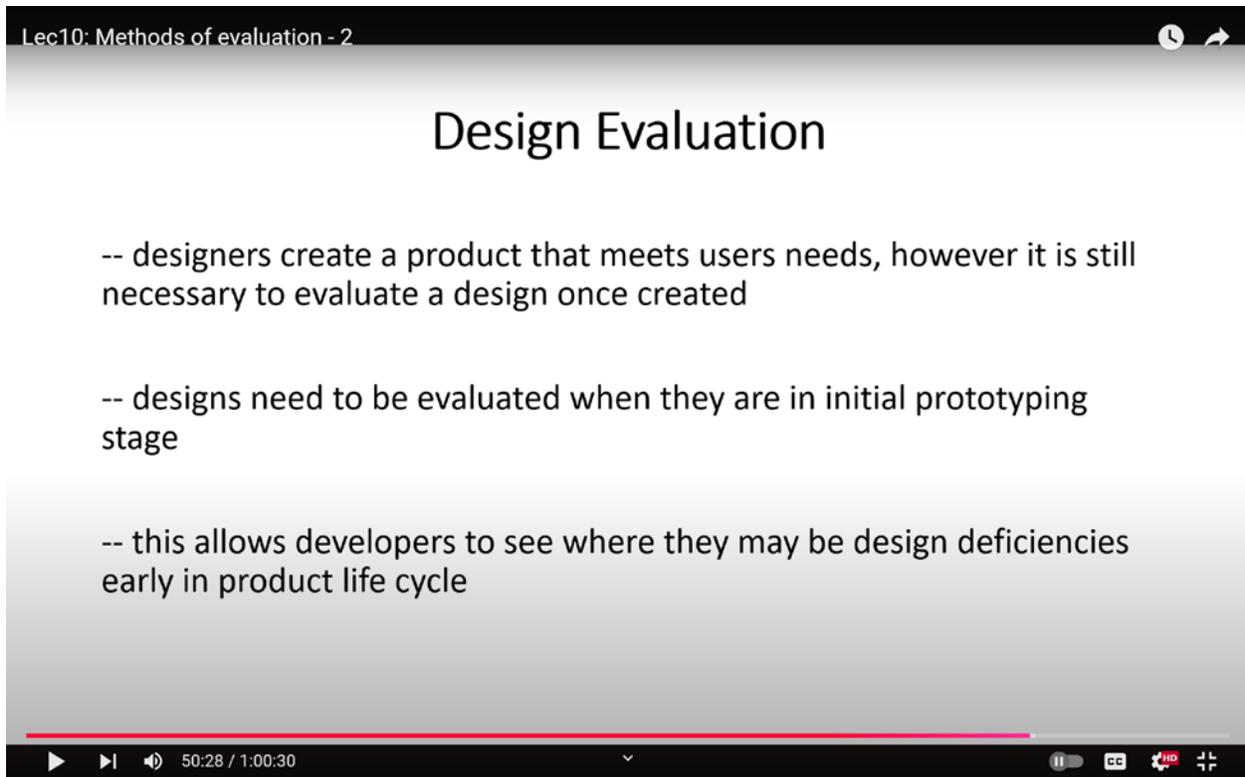
47:21 / 1:00:30

One method used in task analysis or design evaluation is card sorting. Card sorting is a technique employed to understand how people categorize information. When a researcher aims to uncover individuals' mental models regarding a specific application or task, card sorting can be an effective approach. This method is beneficial because it integrates data from various collection methods, resulting in a rich dataset that provides sufficient quality information for making design modifications.

Now, researchers present target users with a series of index cards, each containing the name of an

object or a piece of information. This activity may also be conducted electronically using virtual cards. During this process, target users are invited to participate. When data is gathered, it is grouped into several categories. These categories are then provided to the target users along with the corresponding index cards.

(Refer Slide Time: 50:28)



Lec10: Methods of evaluation - 2

## Design Evaluation

- designers create a product that meets users needs, however it is still necessary to evaluate a design once created
- designs need to be evaluated when they are in initial prototyping stage
- this allows developers to see where they may be design deficiencies early in product life cycle

50:28 / 1:00:30

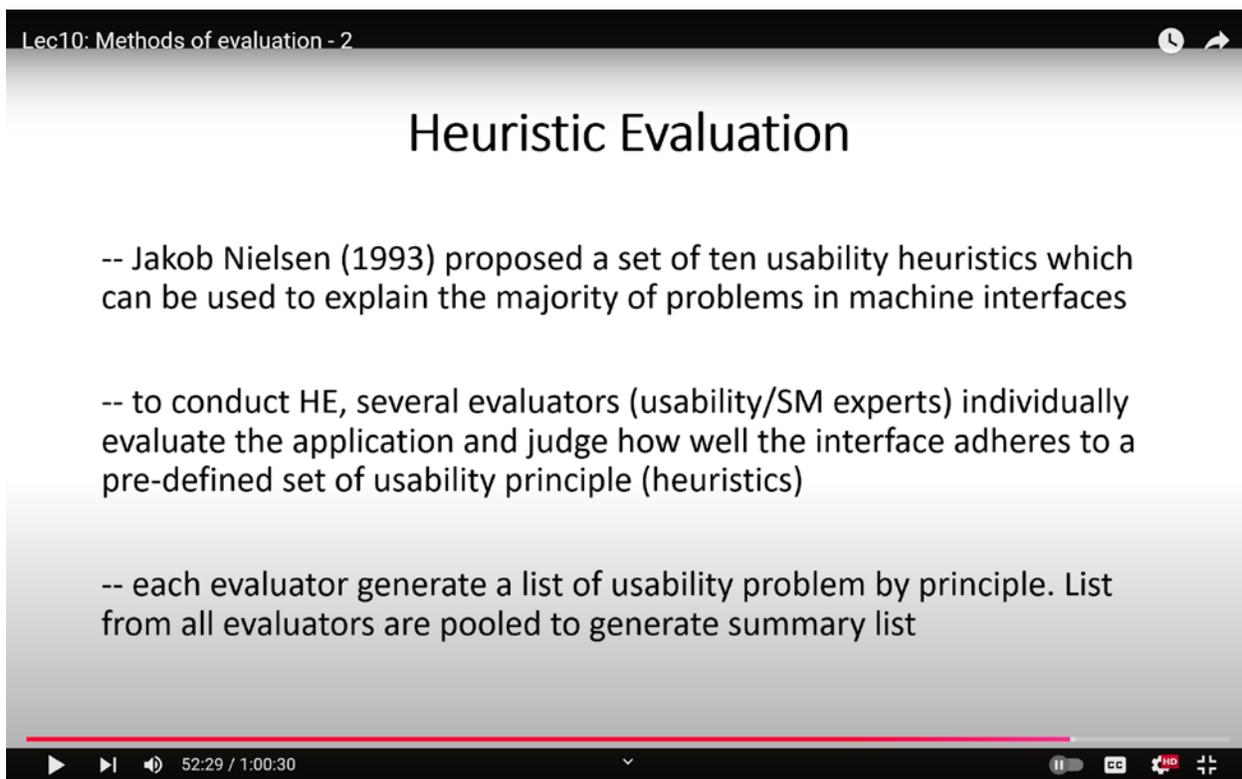
The task for the users is to group together those characteristics and pieces of information that they perceive as similar. By employing card sorting, developers can gain insights into the mental models of the users. There may be a discrepancy between how the developer conceptualizes the information and how the user interprets it. Through card sorting, developers can understand how users categorize various features or pieces of information obtained through different data collection techniques. Users are asked to organize or categorize the cards in a way that makes the most sense to them.

It is possible that the developer has a different mental model of the information than the users do. By utilizing card sorting, the differences between the developer's understanding and the user's

comprehension can be identified. Creating a bridge between the mental models of the user and the developer can lead to a more efficient design. The results of card sorting are generated for each user group to determine any differences in categorization.

The final section of this lecture focuses on design evaluation. After gathering information, studying the user, and collecting data through various methods, it is essential to perform a task analysis. Based on this analysis, modifications can be made, but the effectiveness of these modifications can only be assessed through design evaluation. Designers create products to meet users' needs; however, evaluating the designs they create is crucial.

(Refer Slide Time: 52:29)



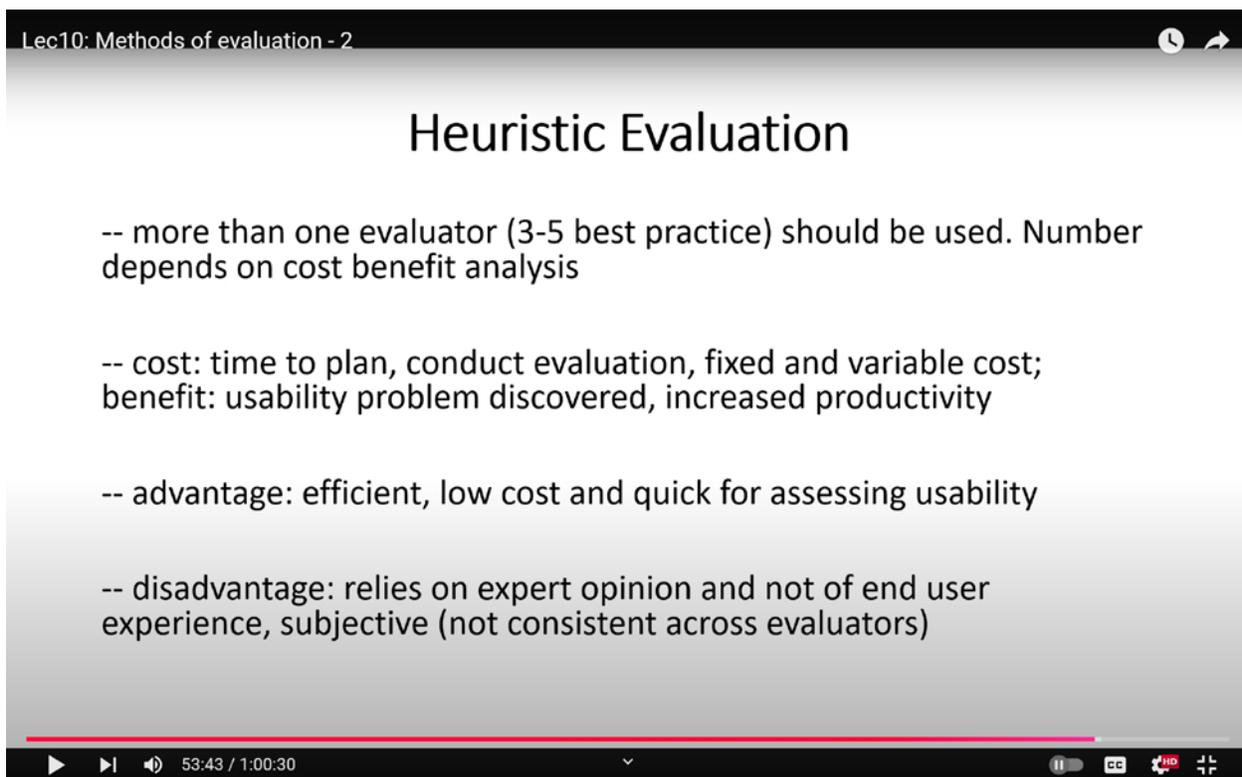
The screenshot shows a video player interface. At the top left, the text 'Lec10: Methods of evaluation - 2' is visible. The main title of the slide is 'Heuristic Evaluation'. Below the title, there are three bullet points: '-- Jakob Nielsen (1993) proposed a set of ten usability heuristics which can be used to explain the majority of problems in machine interfaces', '-- to conduct HE, several evaluators (usability/SM experts) individually evaluate the application and judge how well the interface adheres to a pre-defined set of usability principle (heuristics)', and '-- each evaluator generate a list of usability problem by principle. List from all evaluators are pooled to generate summary list'. At the bottom of the video player, there is a progress bar and control icons, with the time '52:29 / 1:00:30' displayed.

Design evaluation is particularly important during the initial prototyping stage. If design analysis is neglected at this early stage, the result may be a product that is ineffective and inefficient. By involving users during the prototyping phase, while the product is still in development, new modifications can be created and tested. This proactive approach helps ensure that when the product is ultimately launched, it does not contain numerous fundamental issues. Design

evaluation enables developers to identify potential deficiencies early in the product life cycle.

There are two primary methods for conducting design evaluation. One is called heuristic evaluation, a technique proposed by Jakob Nielsen. Nielsen developed a set of ten usability heuristics that evaluators can use to identify major problems in machine interfaces. He provided a list of ten different heuristics that evaluators must follow while assessing a design interface, requiring them to report any errors they find and their relation to specific heuristics. By collecting data from multiple evaluators, developers can identify which heuristic criteria the design fails to meet, allowing for targeted improvements.

(Refer Slide Time: 53:43)



Lec10: Methods of evaluation - 2

## Heuristic Evaluation

- more than one evaluator (3-5 best practice) should be used. Number depends on cost benefit analysis
- cost: time to plan, conduct evaluation, fixed and variable cost; benefit: usability problem discovered, increased productivity
- advantage: efficient, low cost and quick for assessing usability
- disadvantage: relies on expert opinion and not of end user experience, subjective (not consistent across evaluators)

53:43 / 1:00:30

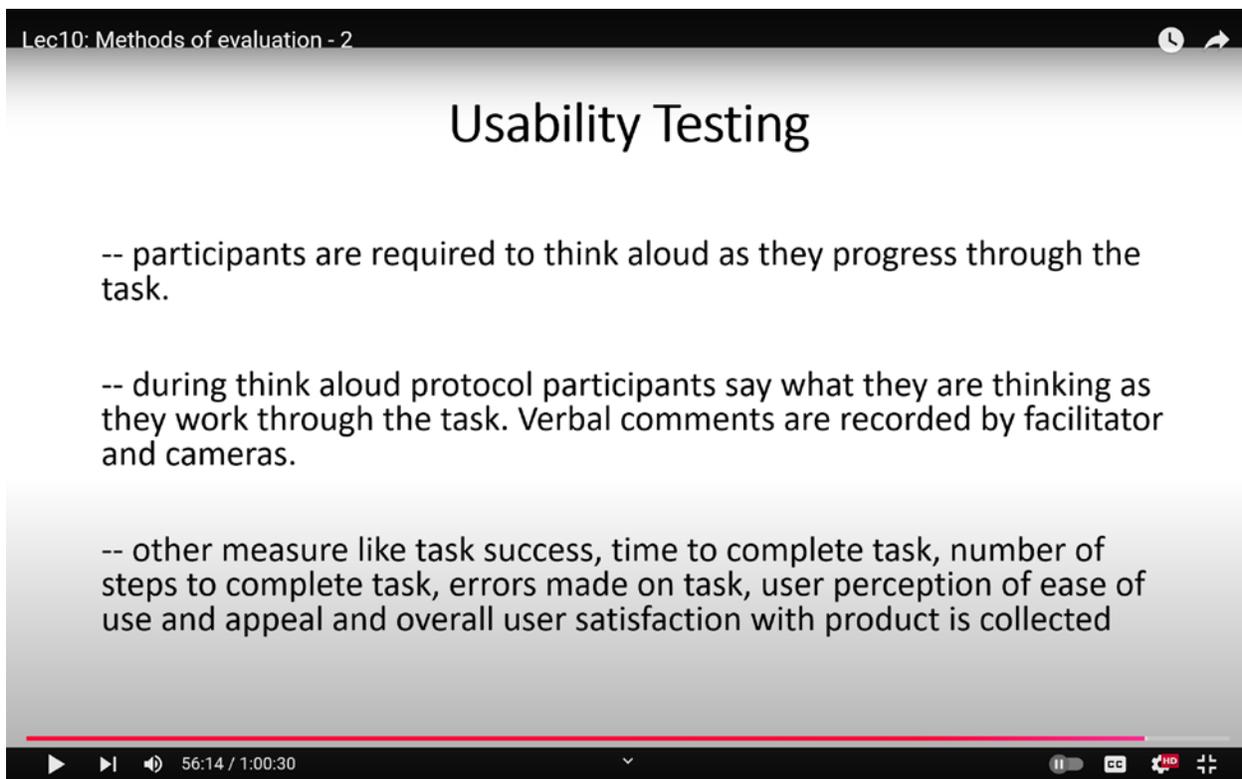
To conduct a heuristic analysis, several evaluators, typically usability experts or subject matter experts, are engaged. Each evaluator independently assesses the application and judges how well the interface adheres to a predefined set of usability principles. Each evaluator generates a list of usability issues organized by principle. By considering the specific tasks, they evaluate each task based on the ten-point scale or the ten heuristics proposed by Nielsen. Problems are documented

along with the relevant heuristic they violate. Once this data is compiled, developers can refer back to the heuristics to see what each measures and implement necessary changes.

All the data from the evaluators is aggregated, resulting in a summary sheet. It is recommended to involve more than one evaluator, generally, three to five is considered best practice. The number of evaluators included in a heuristic analysis is determined by a cost-benefit analysis. Costs include the time required to plan and conduct the evaluation, as well as fixed and variable expenses such as salaries and benefits. The benefits encompass the usability problems discovered and the resulting increase in productivity.

The advantage of heuristic analysis lies in its efficiency, low cost, and quick assessment of usability. However, a significant disadvantage is that it relies on expert opinions rather than direct feedback from end users. This reliance on experts can lead to issues, as their mental models and approaches to tasks may differ from those of actual users. Moreover, the analysis can be subjective, with different experts providing varying evaluations or definitions of the same feature.

(Refer Slide Time: 56:14)



Lec10: Methods of evaluation - 2

## Usability Testing

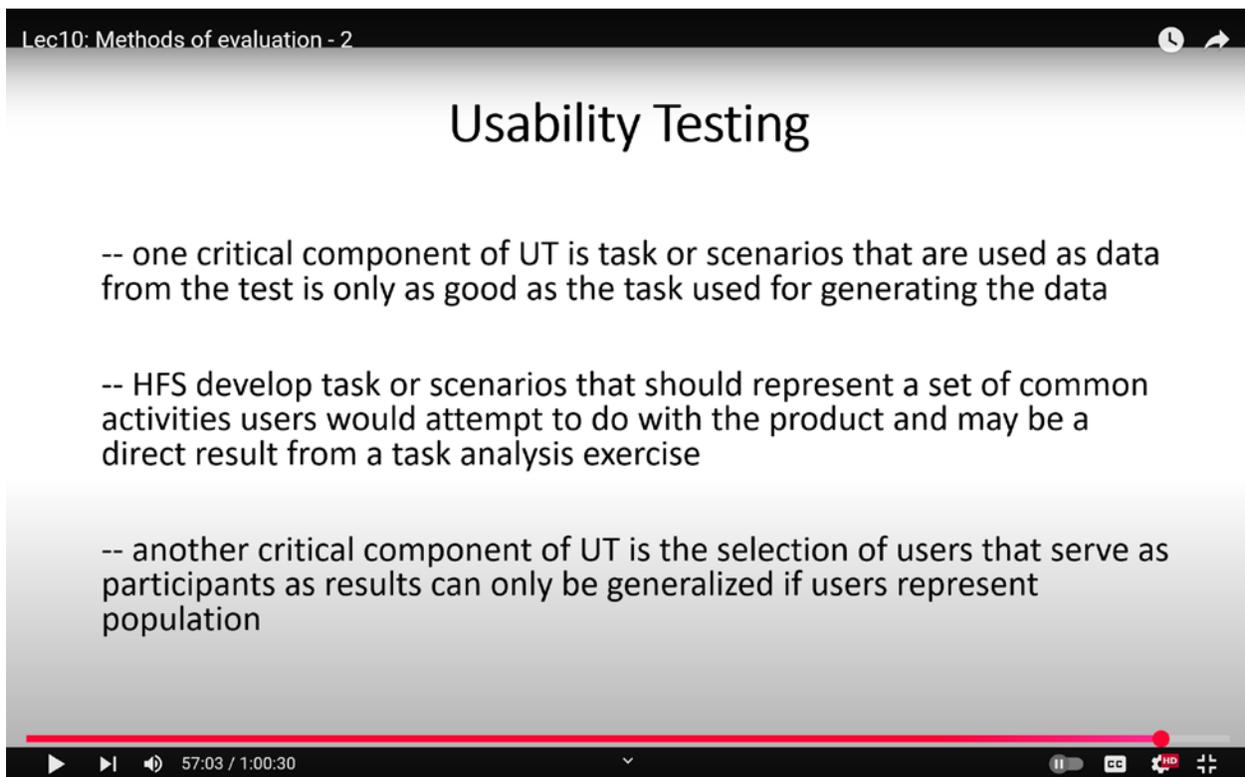
- participants are required to think aloud as they progress through the task.
- during think aloud protocol participants say what they are thinking as they work through the task. Verbal comments are recorded by facilitator and cameras.
- other measure like task success, time to complete task, number of steps to complete task, errors made on task, user perception of ease of use and appeal and overall user satisfaction with product is collected

56:14 / 1:00:30

The ten heuristics proposed by Nielsen should be referenced during evaluations, ensuring that each assessment aligns with these markers. Another method of design evaluation is usability testing. Usability testing is an empirical approach to measuring a product's ease of use. It focuses on how intuitive and accessible a product is for users.

During usability testing, individuals representing the target user group are invited into a usability lab and asked to complete a series of tasks with the product. Researchers observe user behavior, focusing on user satisfaction and performance data, which are subsequently collected and summarized for review. Each participant works one-on-one with a facilitator whose role is to ensure the user completes the tasks correctly without providing guidance on how to perform them. While users engage with the product, they are encouraged to think aloud, articulating their thought processes step by step as they navigate the tasks.

(Refer Slide Time: 57:03)



Lec10: Methods of evaluation - 2

## Usability Testing

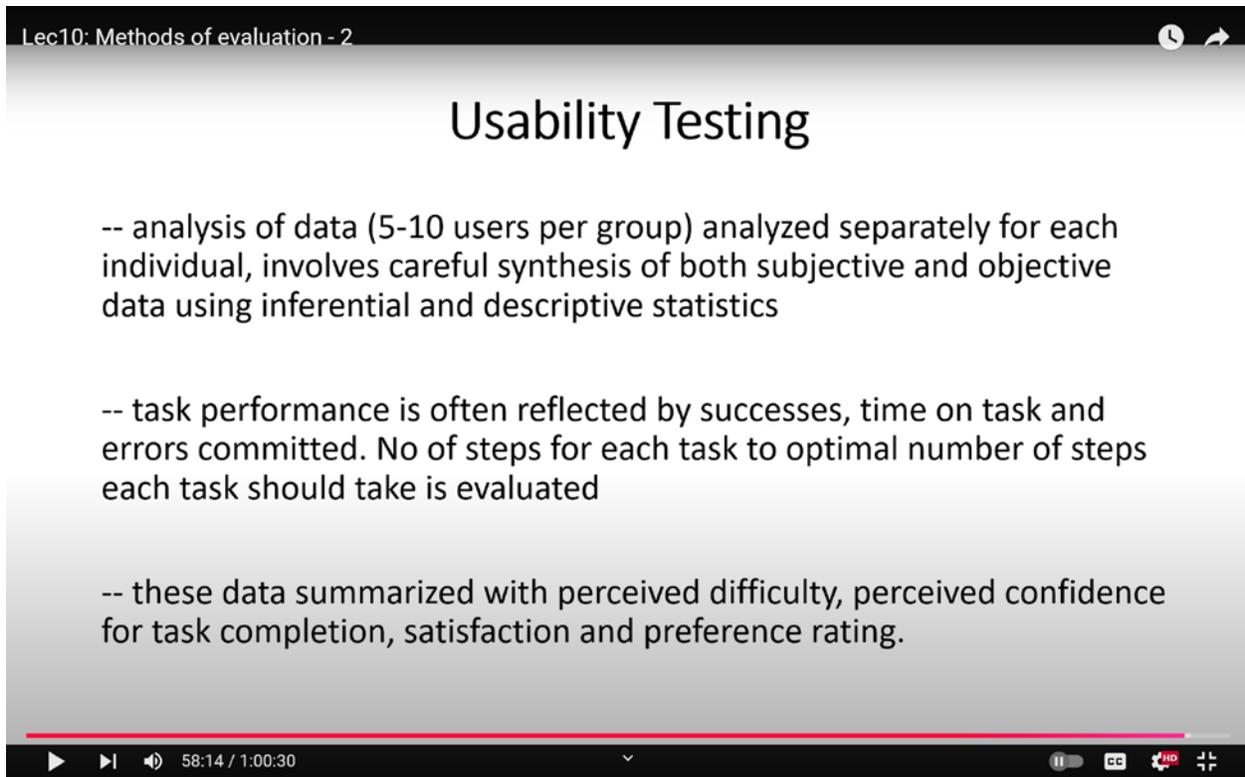
- one critical component of UT is task or scenarios that are used as data from the test is only as good as the task used for generating the data
- HFS develop task or scenarios that should represent a set of common activities users would attempt to do with the product and may be a direct result from a task analysis exercise
- another critical component of UT is the selection of users that serve as participants as results can only be generalized if users represent population

57:03 / 1:00:30

The think-aloud protocol allows participants to verbalize their thoughts during the task, and these verbal comments are recorded by facilitators and cameras. Additional metrics collected during

usability testing include task success rates, the time taken to complete tasks, the number of steps required to accomplish a task, errors made, users' perceptions of ease of use and appeal, and overall satisfaction with the product. This comprehensive data collection informs the evaluation of the product's usability.

(Refer Slide Time: 58:14)



Lec10: Methods of evaluation - 2

## Usability Testing

- analysis of data (5-10 users per group) analyzed separately for each individual, involves careful synthesis of both subjective and objective data using inferential and descriptive statistics
- task performance is often reflected by successes, time on task and errors committed. No of steps for each task to optimal number of steps each task should take is evaluated
- these data summarized with perceived difficulty, perceived confidence for task completion, satisfaction and preference rating.

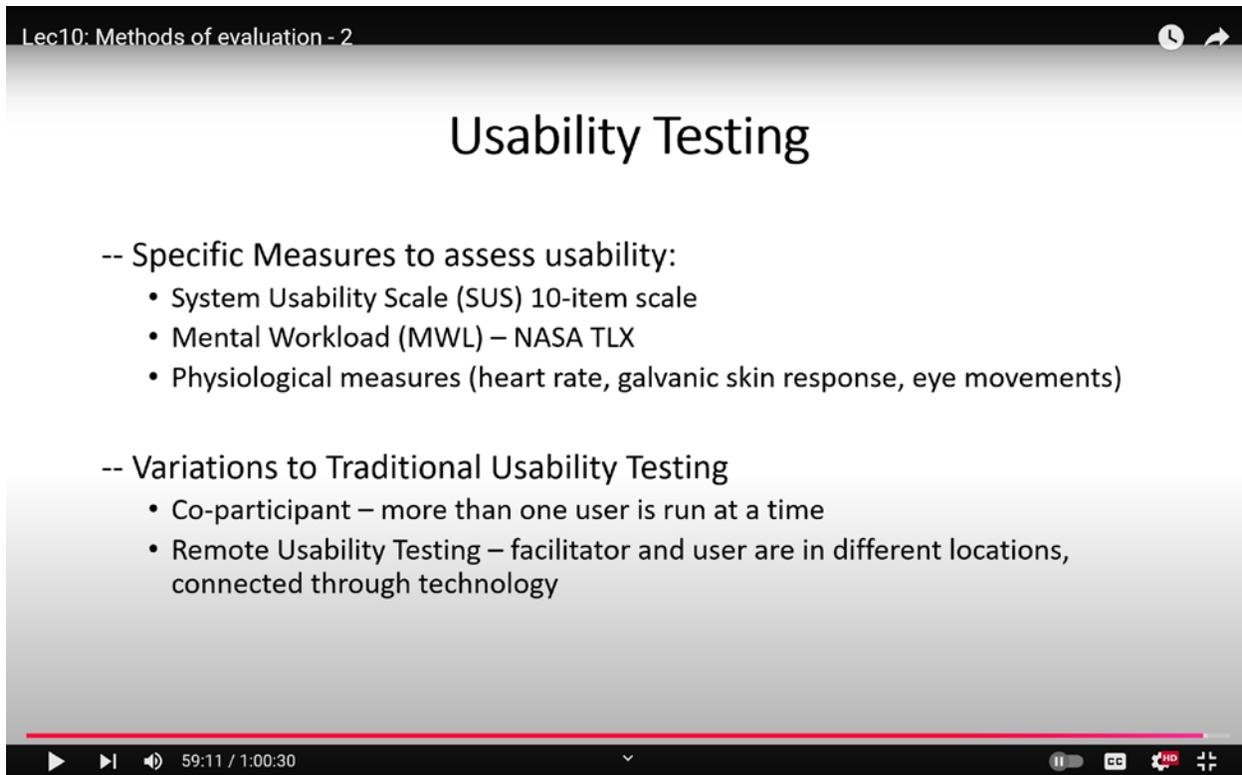
58:14 / 1:00:30

Now, one critical component of usability testing is the tasks or scenarios that are utilized as data for the tests, as the quality of a test is directly linked to the tasks employed. Various scenarios can be created to perform the same task and collect data from it. It is essential that the same scenario is not used for every user, as different users operate in distinct environments. Therefore, a range of hypothetical scenarios is developed, and users are tested across these different scenarios. Human factor specialists design tasks or scenarios that represent a set of common activities that users would typically attempt to accomplish with the product, which may arise from a task analysis exercise.

Another vital aspect of usability testing is the selection of users who participate in the study.

Results can be generalized only if the participants accurately represent the target population. If the selected users are varied or not part of the intended user base for the product, the data collected will likely be inaccurate. Thus, it is crucial to include users who are expected to use the product or are engaged with it.

(Refer Slide Time: 59:11)



Lec10: Methods of evaluation - 2

## Usability Testing

- Specific Measures to assess usability:
  - System Usability Scale (SUS) 10-item scale
  - Mental Workload (MWL) – NASA TLX
  - Physiological measures (heart rate, galvanic skin response, eye movements)
- Variations to Traditional Usability Testing
  - Co-participant – more than one user is run at a time
  - Remote Usability Testing – facilitator and user are in different locations, connected through technology

59:11 / 1:00:30

Data analysis is performed, generally involving groups of 5 to 10 participants. Each individual's data is analyzed separately. This analysis encompasses both subjective and objective data analysis methods, employing both inferential and descriptive statistics. Task performance is typically reflected in measures such as success rates, the time spent on tasks, and errors committed. The number of steps required for each task is compared to the optimal number of steps, which provides insight into how the task is performed and what the workflow entails.

The data collected are summarized with metrics such as perceived difficulty, perceived confidence in task completion, satisfaction levels, and preference ratings. This process forms the basis of the evaluation. Various methods can be employed in usability testing, and specific measures to assess

usability include the System Usability Scale (SUS), which consists of a 10-item scale. Other tools such as the Mental Workload (MWL) and NASA Task Load Index (TLX), along with questionnaires that address usability-related concepts, can be utilized. Additionally, physiological measures, such as heart rate, galvanic skin response, and eye movements, can also be included.

There are variations of traditional usability testing, including co-participant usability testing, where multiple users are present simultaneously, and remote usability testing, in which the facilitator and the user are in different locations and connected via technology.

In this lecture, we examined how task analysis is conducted and how questionnaires, focus groups, and contextual inquiries provide data for analysis. We explored heuristic evaluation and usability testing as methods of design evaluation. Overall, we gained an understanding of the various evaluation methods for designs and design modifications to ensure that the revised designs fulfill their intended purpose and enhance their efficiency. This concludes our session for today from the MOOCs lab.