Chapter 11 - Introducing Evaluation

1. Usability Testing

Definition: Usability testing is a method that involves evaluating user interfaces by collecting data in a controlled setting, often involving typical tasks and users for whom the system is designed.

When and Why It Is Used: This method is used to ensure that a product is easy to use, effective, and satisfying for end users. It is most commonly used in lab settings during iterative stages of design to refine product usability before release.

Steps/Principles:

- Recruit participants that represent the target users.
- Design and assign tasks typical of user goals.
- Record metrics like time taken, errors made, and user feedback through observation and questionnaires.
- Conduct interviews to gather user satisfaction data.

Origin: Usability testing is an established process in Human-Computer Interaction (HCI) and UX design. Pioneers like Jakob Nielsen in the early 1990s popularized it.

2. In-the-Wild Studies

Definition: This method involves observing how technology is used in natural settings with little or no intervention from evaluators.

When and Why It Is Used: In-the-wild studies are used to evaluate how products are utilized in real-life scenarios. They provide insights into how users naturally interact with technology, revealing unforeseen issues.

Steps/Principles:

- Deploy the prototype in natural settings (homes, public areas).
- Observe and record user interactions, noting behaviors, issues, and context.
- Collect feedback through logs, diaries, or interviews.

Origin: The use of "in-the-wild" studies began emerging as a trend in the 1990s and 2000s to complement the limitations of lab studies in capturing authentic user behavior.

3. Heuristic Evaluation

Definition: Heuristic evaluation is a predictive evaluation method where usability experts assess the interface against established heuristics or usability principles.

When and Why It Is Used: Used when the goal is to quickly identify major usability problems without involving actual users. Suitable during early design stages or for quick analysis.

Steps/Principles:

- Experts evaluate the product, comparing its features against known usability heuristics.
- Identify issues such as visibility, match with user expectations, and error prevention.
- Document and categorize findings for prioritization.

Origin: Jakob Nielsen and Rolf Molich developed heuristic evaluation in the early 1990s, defining ten widely used heuristics.

4. Cognitive Walk-Through

Definition: This evaluation method focuses on simulating a user's cognitive process when interacting with an interface, assessing how easily users can learn and navigate the system.

When and Why It Is Used: It is primarily used for evaluating designs for ease of learning, especially when considering novice users. Suitable during early stages of design.

Steps/Principles:

- Define the tasks to be performed by users.
- Go step-by-step through each task, evaluating the ease of progression.
- Check whether each action would make sense to the user, and identify points where users may struggle.

Origin: Developed by Clayton Lewis, Peter Polson, and others in the 1990s.

5. Controlled Experiments

Definition: Controlled experiments involve testing participants in lab settings where variables such as environment, time, and tasks are manipulated to observe their effects on user interaction.

When and Why It Is Used: Used to measure specific user interactions under controlled conditions, such as comparing two different interface designs to determine which yields better usability results.

Steps/Principles:

- Define a hypothesis to be tested.
- Design controlled conditions to minimize outside influences.
- Measure variables such as task completion time, accuracy, and user preferences.

Origin: Controlled experiments have roots in experimental psychology, adopted for HCI evaluation in the 1980s.

6. Analytics

Definition: Analytics involve collecting and analyzing quantitative data, such as user interactions, click paths, and time spent on pages, to evaluate a system’s performance.

When and Why It Is Used: Used for evaluating already deployed products to understand how users engage with the system, determine issues, and identify areas for improvement.

Steps/Principles:

- Set up logging or analytics software.

- Collect quantitative data, such as bounce rates, conversion rates, or task paths.
- Analyze data to identify bottlenecks or areas for enhancement.

Origin: Analytics became more popular with web and software evaluation during the growth of the Internet in the late 1990s.

7. Remote Evaluation

Definition: This method involves evaluating participants via the internet, either by observing through video calls or logging interactions while the user is in their own environment.

When and Why It Is Used: Remote evaluation is used when it is impractical to bring users to a lab or when a wider, more geographically dispersed user base needs to be evaluated. Became particularly useful during the COVID-19 pandemic.

Steps/Principles:

- Recruit participants remotely and provide them with software or product access.
- Use video conferencing, screen sharing, or automated logging to observe tasks.
- Gather data remotely through questionnaires or follow-up interviews.

Origin: Remote evaluation practices date back to the 1990s with pioneers like Rex Hartson, becoming especially prevalent during the 2020s.

8. Crowdsourcing

Definition: Crowdsourcing involves collecting data or contributions from a large group of people, often remotely over the internet, to evaluate a product, complete a survey, or provide feedback.

When and Why It Is Used: Used when a large volume of data or rapid feedback is needed. It can also help reach a diverse participant pool to generalize findings effectively.

Steps/Principles:

- Post tasks to crowdsourcing platforms (e.g., Amazon Mechanical Turk).
- Define specific tasks (e.g., rating designs, tagging images, completing user tests).
- Collect and analyze data, often across a broad and diverse population.

Origin: Jeff Heer and Michael Bostock first used crowdsourcing for HCI studies in the early 2010s, demonstrating its efficiency in gaining quick, diverse insights.