Prof. Joseph K Bada and Brian Lugemwa

Human Computer Interaction User Interface Evaluation

Presentation Outline

- Introduction

- The Why, What, Where, and When of Evaluation

- Types of Evaluation

- Evaluating for Accessibility Using the Web Content Accessibility Guidelines

- Evaluation—Inspections, Analytics, and Models

Introduction

- *Evaluation* is integral to the design process. It involves collecting and analyzing data about users' or potential users' experiences when interacting with a design artifact such as a screen sketch, prototype, app, computer system, or component of a computer system.

- A central goal of evaluation is to improve the artifact's design. Evaluation focuses on both the usability of the system (that is, how easy it is to learn and to use) and on the users' experiences when interacting with it (for example, how satisfying, enjoyable, or motivating the interaction is).

Introduction (continued)

- Devices such as smartphones, iPads, and e-readers, together with the pervasiveness of mobile apps and the emergence of IoT devices, have heightened awareness about usability and interaction design.

- However, many designers still assume that if they and their colleagues can use a product and find it attractive, others will too.

- The problem with this assumption is that designers may then design only for themselves. Evaluation enables them to check that their design is appropriate and acceptable for the target user population.

Introduction (continued)

There are many different evaluation methods. Which to use depends on the goals of the evaluation.

Evaluations can occur in a range of places such as in labs, people's homes, outdoors, and work settings.

Evaluations usually involve observing participants and measuring their performance during usability testing, experiments, or field studies in order to evaluate the design or design concept.

There are other methods, however, that do not involve participants directly, such as modeling users' behavior and analytics.

Modeling users' behavior provides an approximation of what users might do when interacting with an interface; these models are often done as a quick way of assessing the potential of different interface configurations.

Analytics provide a way of examining the performance of an already existing product, such as a website, so that it can be improved.

Why Evaluate?

- User experience involves all aspects of the user's interaction with the product. Nowadays users expect much more than just a usable system—they also look for a pleasing and engaging experience from more products. Simplicity and elegance are valued so that the product is a joy to own and use.

- From a business and marketing perspective, well-designed products sell. Hence, there are good reasons for companies to invest in evaluating the design of products.

- Designers can focus on real problems and the needs of different user groups and make informed decisions about the design, rather than on debating what each other likes or dislikes. It also enables problems to be fixed before the product goes on sale.

What to Evaluate

- What to evaluate ranges from low-tech prototypes to complete systems, from a particular screen function to the whole workflow, and from aesthetic design to safety features.

- Developers of a new web browser may want to know whether users find items faster with their product.

- Developers of an ambient display may be interested in whether it changes people's behavior.

- Game app developers will want to know how engaging and fun their games are compared with those of their competitors and how long users will play them.

- Government authorities may ask if a computerized system for controlling traffic lights results in fewer accidents or if a website complies with the standards required for users with disabilities.

- Makers of a toy may ask whether 6-year-olds can manipulate the controls, whether they are engaged by its furry cover, and whether the toy is safe for children.

Where to Evaluate

- Where evaluation takes place depends on what is being evaluated. Some characteristics, such as web accessibility, are generally evaluated in a lab because it provides the control necessary to investigate systematically whether all of the requirements are met. This is also true for design choices, such as choosing the size and layout of keys for a small handheld device for playing games.

- Remote studies of online behavior, such as social networking, can be conducted to evaluate natural interactions of participants in the context of their interaction, for example, in their own homes or place of work.

When to Evaluate

- The stage in the product lifecycle when evaluation takes place depends on the type of product and the development process being followed. For example, the product being developed could be a new concept, or it could be an upgrade to an existing product.

- It could also be a product in a rapidly changing market that needs to be evaluated to see how well the design meets current and predicted market needs. If the product is new, then considerable time is usually invested in market research and discovering user requirements.

- Once these requirements have been established, they are used to create initial sketches, a storyboard, a series of screens, or a prototype of the design ideas.

- These are then evaluated to see whether the designers have interpreted the users' requirements correctly and embodied them in their designs appropriately. The designs will be modified according to the evaluation feedback and new prototypes developed and subsequently evaluated.

When to Evaluate (continued)

- When evaluations are conducted during design to check that a product continues to meet users' needs, they are known as *formative evaluations*.

- Formative evaluations cover a broad range of design processes, from the development of early sketches and prototypes through to tweaking and perfecting a nearly finished design.

- Evaluations that are carried out to assess the success of a finished product are known as *summative evaluations*.

- If the product is being upgraded, then the evaluation may not focus on discovering new requirements but may instead evaluate the existing product to ascertain what needs improving.

- Features are then often added, which can result in new usability problems. At other times, attention is focused on improving specific aspects, such as enhanced navigation.

Types of Evaluation

We classify evaluations into three broad categories, depending on the setting, user involvement, and level of control. These are as follows:

- *Controlled settings directly involving users* (examples are usability labs and research labs): Users' activities are controlled to test hypotheses and measure or observe certain behaviors. The main methods are usability testing and experiments.

- *Natural settings involving users* (examples are online communities and products that are used in public places): There is little or no control of users' activities to determine how the product would be used in the real world. The main method used is field studies.

- *Any settings not directly involving users*: Consultants and researchers critique, predict, and model aspects of the interface to identify the most obvious usability problems. The range of methods includes inspections, heuristics, walk-throughs, models, and analytics.

Controlled Settings Involving Users

- Experiments and user tests are designed to control what users do, when they do it, and for how long. They are designed to reduce outside influences and distractions that might affect the results, such as people talking in the background.

- The approach has been extensively and successfully used to evaluate software applications running on laptops and other devices where participants can be seated in front of them to perform a set of tasks.

Usability Testing

- This approach to evaluating user interfaces involves collecting data using a combination of methods in a controlled setting, for example, experiments that follow basic experimental design, observation, interviews, and questionnaires.

- Often, *usability testing* is conducted in labs, although increasingly interviews and other forms of data collection are being done remotely via phone and digital communication (for instance, through Skype or Zoom) or in natural settings.

Controlled Settings Involving Users (continued)

- The primary goal is to determine whether an interface is usable by the intended user population to carry out the tasks for which it was designed. This involves investigating how typical users perform on typical tasks.

- By typical, we mean the users for whom the system is designed (for example, teenagers, adults, and so on) and the activities that it is designed for them to be able to do (such as, purchasing the latest fashions).

- It often involves comparing the number and kinds of errors that users make between versions and recording the time that it takes them to complete the task. As users perform the tasks, they may be recorded on video.

- Their interactions with the software may also be recorded, usually by logging software.

- User satisfaction questionnaires and interviews can also be used to elicit users' opinions about how they liked the experience of using the system.

- This data can be supplemented by observation at product sites to collect evidence about how the product is being used in the workplace or in other environments.

Natural Settings Involving Users

- The goal of field studies is to evaluate products with users in their natural settings. Field studies are used primarily to

  - Help identify opportunities for new technology

  - Establish the requirements for a new design

  - Facilitate the introduction of technology or inform deployment of existing technology in new contexts

- Methods that are typically used are observation, interviews, and interaction logging.

- The data takes the form of events and conversations that are recorded by the researchers as notes, or through audio or video recording, or by the participants as diaries and notes.

- The goal is to be unobtrusive and not to affect what people do during the evaluation. However, it is inevitable that some methods will influence how people behave.

- For example, diary studies require people to document their activities or feelings at certain times, and this can make them reflect on and possibly change their behavior.

Any Settings Not Involving Users

- Evaluations that take place without involving users are conducted in settings where the researcher has to imagine or model how an interface is likely to be used.

- Inspection methods are commonly employed to predict user behavior and to identify usability problems based on knowledge of usability, users' behavior, the contexts in which the system will be used, and the kinds of activities that users undertake.

- Examples include heuristic evaluation that applies knowledge of typical users guided by rules of thumb and walkthroughs that involve stepping through a scenario or answering a set of questions for a detailed prototype. Other techniques include analytics and models.

Any Settings Not Involving Users (continued)

- The original heuristics used in heuristic evaluation were for screen-based applications. These have been adapted to develop new sets of heuristics for evaluating web-based products, mobile systems, collaborative technologies, computerized toys, information visualizations and other new types of systems.

- *Cognitive walk-throughs* involve simulating a user's problem-solving process at each step in the human-computer dialogue and checking to see how users progress from step to step in these interactions.

- *Analytics* is a technique for logging and analyzing data either at a customer's site or remotely.

- *Web analytics* is the measurement, collection, analysis, and reporting of Internet data to understand and optimize web usage.

- Examples of web analytics include the number of visitors to a website home page over a particular time period, the average time users spend on the home page, which other pages they visit, or whether they leave after visiting the homepage.

Evaluation—Inspections, Analytics, and Models

- None of these methods requires users to be present during the evaluation. Inspection methods often involve a researcher, sometimes known as an *expert*, role-playing the users for whom the product is designed, analyzing aspects of an interface, and identifying potential usability problems.

- The most well-known methods are *heuristic evaluation* and *walk-throughs*.

- *Analytics* involves user interaction logging, and *A/B testing* is an experimental method. Both analytics and A/B testing are usually carried out remotely.

- *Predictive modeling* involves analyzing the various physical and mental operations that are needed to perform particular tasks at the interface and operationalizing them as quantitative measures.

Inspections: Heuristic Evaluation and Walk-Throughs

- Sometimes, it is not practical to involve users in an evaluation because they are not available, there is insufficient time, or it is difficult to find people. In such circumstances, other people, often referred to as *experts* or *researchers*, can provide feedback. These are people who are knowledgeable about both interaction design and the needs and typical behavior of users.

- Inspection methods for interaction design include heuristic evaluations and walk-throughs, in which researchers examine the interface of an interactive product, often role-playing typical users, and suggest problems that users would likely have when interacting with the product.

- One of the attractions of these methods is that they can be used at any stage of a design project. They can also be used to complement user testing.

Heuristic Evaluation

- In *heuristic evaluation*, researchers, guided by a set of usability principles known as *heuristics*, evaluate whether user-interface elements, such as dialog boxes, menus, navigation structure, online help, and so on, conform to tried-and-tested principles. These heuristics closely resemble high-level design principles.

- The set of heuristics for HCI evaluation are presented below:

  - **Visibility of System Status** The system should always keep users informed about what is going on, through appropriate feedback and within reasonable time.

  - **Match Between System and the Real World** The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. It should follow real-world conventions, making information appear in a natural and logical order.

  - **User Control and Freedom** Users often choose system functions by mistake and will need a clearly marked emergency exit to leave the unwanted state without having to go through an extended dialog. The system should support undo and redo.

  - **Consistency and Standards** Users should not have to wonder whether different words, situations, or actions mean the same thing. The system should follow platform conventions.

Heuristic Evaluation (continued)

- **Error Prevention** Rather than just good error messages, the system should incorporate careful design that prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

- **Recognition** Rather Than Recall Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialog to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

- **Flexibility and Efficiency of Use** Accelerators—unseen by the novice user—may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

- **Aesthetic and Minimalist Design** Dialogs should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialog competes with the relevant units of information and diminishes their relative visibility.

- **Help Users Recognize, Diagnose, and Recover from Errors** Error messages should be expressed in plain language (not codes), precisely indicate the problem, and constructively suggest a solution.

- **Help and Documentation** Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Heuristic Evaluation for Websites

Clarity

- Make the system as clear, concise, and meaningful as possible for the intended audience.
  - Write clear, concise copy.
  - Only use technical language for a technical audience.
  - Write clear and meaningful labels.
  - Use meaningful icons.

Minimize Unnecessary Complexity and Cognitive Load

- Make the system as simple as possible for users to accomplish their tasks.
  - Remove unnecessary functionality, process steps, and visual clutter.
  - Use progressive disclosure to hide advanced features.
  - Break down complicated processes into multiple steps.
  - Prioritize using size, shape, color, alignment, and proximity.

Heuristic Evaluation for Websites (continued)

Provide Users with Context

- Interfaces should provide users with a sense of context in time and space.

  - Provide a clear site name and purpose.

  - Highlight the current section in the navigation.

  - Provide a breadcrumb trail (that is, show where the user has been in a website).

  - Use appropriate feedback messages.

  - Show the number of steps in a process.

  - Reduce perception of latency by providing visual cues (for instance, a progress indicator) or by allowing users to complete other tasks while waiting.

Heuristic Evaluation for Websites (continued)

Promote a Pleasurable and Positive User Experience

• The user should be treated with respect, and the design should be aesthetically pleasing and promote a pleasurable and rewarding experience.

- Create a pleasurable and attractive design.

- Provide easily attainable goals.

- Provide rewards for usage and progression.

Evaluating for Accessibility Using the Web Content Accessibility Guidelines

- *Web Content Accessibility Guidelines (WCAG)* are a detailed set of standards about how to ensure that web page content is accessible for users with various disabilities.

- Heuristic evaluation are well-known within the HCI community, the WCAG is probably the best-known set of interface guidelines or standards outside of the HCI community.

- Why? Because many countries around the world have laws that require that government websites, and websites of public accommodations (such as hotels, libraries, and retail stores), are accessible for people with disabilities.

Evaluating for accessibility (continued)

The key concepts of web accessibility, according to WCAG, are summarized as POUR—Perceivable, Operable, Understandable, and Robust.

- **Perceivable**
  - 1.1 Provide text alternatives for non-text content.
  - 1.2 Provide captions and other alternatives for multimedia.
  - 1.3 Create content that can be presented in different ways, including by assistive technologies, without losing meaning.
  - 1.4 Make it easier for users to see and hear content.

Evaluating for accessibility (continued)

- **Operable**
  - 2.1 Make all functionality available from a keyboard.
  - 2.2 Give users enough time to read and use content.
  - 2.3 Do not use content that causes seizures or physical reactions.
  - 2.4 Help users navigate and find content.
  - 2.5 Make it easy to use inputs other than keyboard.
- **Understandable**
  - 3.1 Make text readable and understandable.
  - 3.2 Make content appear and operate in predictable ways.
  - 3.3 Help users avoid and correct mistakes.
- **Robust**
  - 4.1 Maximize compatibility with current and future user tools.

Walk-Throughs

• Walk-throughs offer an alternative approach to heuristic evaluation for predicting users' problems without doing user testing. As the name suggests, **walk-throughs** involve walking through a task with the product and noting problematic usability features.

• While most walk-through methods do not involve users, others, such as pluralistic walk-throughs, involve a team that may include users, as well as developers and usability specialists.

• **Cognitive walk-throughs** involve simulating how users go about problem-solving at each step in a human-computer interaction.

• A cognitive walk-through, as the name implies, takes a cognitive perspective in which the focus is on evaluating designs for ease of learning—a focus that is motivated by observations that users learn by exploration.

Analytics and A/B Testing

- A variety of users' actions can be recorded by software automatically, including key presses, mouse or other pointing device movements, time spent searching a web page, looking at help systems, and task flow through software modules.

- A key advantage of logging activity automatically is that it is unobtrusive provided the system's performance is not affected, but it also raises ethical concerns about observing participants if this is done without their knowledge.

- Another advantage is that large volumes of data can be logged automatically and then explored and analyzed using visualization and other tools.

Web Analytics

- *Web analytics* is a form of interaction logging that was specifically created to analyze users' activity on websites so that designers could modify their designs to attract and retain customers.

- Using web analytics, web designers and developers can trace the activity of the users who visit their website.

- They can see how many people came to the site, how many stayed and for how long, and which pages they visited.

- They can also find out about where the users came from and much more.

- Web analytics is therefore a powerful evaluation tool for web designers that can be used on its own or in conjunction with other types of evaluations, particularly user testing. For instance, web analytics can provide a "big-picture" overview of user interaction on a website.

A/B Testing

- Another way to evaluate a website, part of a website, an application, or an app running on a mobile device is by carrying out a large-scale experiment to evaluate how two groups of users perform using two different designs— one of which acts as the control and the other as the experimental condition, that is, the new design being tested.

- This approach is known as *A/B testing*, and it is basically a controlled experiment but one that often involves hundreds or thousands of participants.

- To do A/B testing, a variable of interest is identified, such as the design of an advertisement. Group A is served with design A, the existing design, and group B is served with design B, the new design.

- A dependent measure is then identified, such as how many times participants in each group, A and B, click the advertisement that they are presented over a particular period of time, such as a day, week, or a month.

Key Points

- Inspections can be used for evaluating a range of representations including requirements, mockups, prototypes, or products.

- User testing and heuristic evaluation often reveal different usability problems.

- Other types of inspections used in UX design include pluralistic and cognitive walk-throughs.

- Walk-throughs are a fine-grained, focused methods that are suitable for evaluating small parts of a product.

- Analytics involve collecting data about user interactions to identify how users use a website or product and which parts are underused.

- When applied to websites, analytics are often referred to as *web analytics*. Similarly, when applied to learning systems, they are referred to as *learning analytics*.

*****THE END*****