User Interface Design

Lecture 8: Usability Evaluation

Dr. Obead Alhadreti

Learning Objectives

- Describe usability evaluation
- Explain usability evaluation types
- Describe usability evaluation methods, and they are conducted.

Usability Evaluation

- Usability evaluation is:

  "the appraisal of a particular application's user interface, an interaction metaphor or method, or an input device, for the reason of ascertaining of determining its real or likely usability" (Koutsabasis et al., 2007)

- Usability evaluation is required at several points during the design process.

Types of Usability Evaluation

- Usability evaluation methods can be classified into expert-based methods, user-based methods, and model-based methods.

1. **Expert-Based Methods:** Expert-based methods are a set of methods that involve having experts assess the usability of an interface, predicting potential usability problems, and providing recommendations for improvement.

- The two most commonly employed expert-based methods are ***heuristic evaluation* and *cognitive walkthrough*.**

Heuristic Evaluation

- Developed by Nielsen and Molich in 1990, and involves usability experts checking whether the interface conforms to a set of guidelines or principles.
- It is a cheap, fast, and easy method for evaluating systems.
- It can be used in the early stages of a system's development.

Heuristic evaluation

- Jakob Nielsen's Ten Usability Heuristics:

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

Procedure of heuristics evaluation

1. Establish an appropriate list of heuristics.
2. Select your evaluators.
3. Brief your evaluators so they know exactly what they are meant to do and cover during their evaluation.
4. First evaluation phase. The evaluators will separately, use the product freely to gain a feel for the methods of interaction and the scope.
5. Second evaluation phase. The evaluators will separately evaluate the interface against the heuristics.
6. Record problems.
7. Debriefing session. The findings and ratings can then be aggregated after they have all finished the evaluation process.

Evaluator Effect

- The method can lead to *unreliable* results as it depends highly on the evaluators' experience or it is referred to as *'evaluator effect'*.

- Evaluator effect refers to

> *"the observation that individual usability evaluators can identify substantially different sets of usability problems when analysing the same test sessions"* (Hertzum et al., 2014)

Cognitive Walkthrough

- Developed by Lewis in 1994, is based not on a set of guidelines but on a set of realistic task scenarios.
- By following these scenarios, experts attempt to discover the usability problems that users might encounter whilst working with the system.
- Experience shows that many users prefer to learn how to use a system by exploring its functionality hands on, and not after sufficient training or examination of a user's manual. So the checks that are made during the walkthrough ask questions that address this exploratory learning.

Procedure of Cognitive Walkthrough

1. A specification or prototype of the system. It doesn't have to be complete, but it should be fairly detailed.
2. A description of the task the user is to perform on the system. This should be a representative task that most users will want to do.
3. A complete, written list of the actions needed to complete the task with the proposed system.
4. An indication of who the users are and what kind of experience and knowledge the evaluators can assume about them.

Procedure of Cognitive Walkthrough

- Given this information, the evaluators step through the action sequence (identified in item 3) to critique the system and tell a believable story about its usability.

- To do this, the evaluators try to answer the following four questions for each step in the action sequence.
  1. Is the effect of the action the same as the user's goal at that point?
  2. Will users see that the action is available?
  3. Once users have found the correct action, will they know it is the one they need?
  4. After the action is taken, will users understand the feedback they get?

User-based Methods

- Many methods exist for conducting user-based evaluation, such as surveys, interviews, and focus groups.

- Surveys, interviews and focus groups are methods which involve simply asking participants what they *think* of a particular test object, but not whether users can actually *work* with the object.

- Another approach is to conduct **usability test.**

Usability testing

- Barnum (2011, p. 13) has defined usability testing as *"the activity that focuses on observing users working with a product, performing tasks that are real and meaningful to them"*.

- The challenge for usability evaluators, however, is that they can see *what* a user is doing but not *why* they are doing it.

- Behavioural observation, such as *think-aloud protocol,* has been developed in response to this challenge.

Usability testing

- The general idea for think-aloud usability testing is that test participants verbally express their intentions, actions, and frustrations whilst (or shortly after) working with an interactive system.
- The usability practitioner then uses this information to identify problem areas of the system being assessed, and to offer recommendations for improvement.
- The main drawback to the think-aloud method is that it can be time-consuming and expensive compared to expert-based or model-based evaluation methods.

Procedure of Usability testing

- Dumas and Redish set out five specific requirements for usability testing:

1. A clear goal;
2. Real or representative users;
3. Real tasks;
4. Observation and recording; and
5. Analysing data and making suggestions for improvements.

Goal of Usability test

- The primary goal of a usability test is to derive a list of usability problems from evaluators' observations and analyses of users' verbal as well as non-verbal behaviour.
- Usability testing may also involve other metrics that seek to gauge usability by measuring performance and/or preference.
- Performance measures (e.g. time spent on tasks, or number of tasks completed successfully) indicate a user's level of capability with the system.
- Preference measures indicate how much the users enjoy using the system.

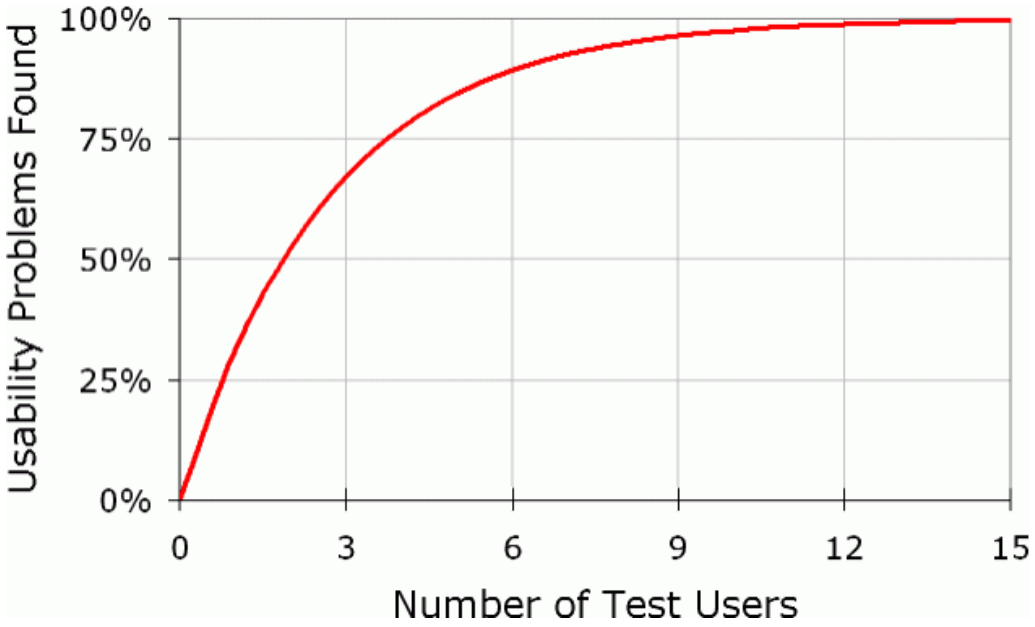
Test Participants

- In a think-aloud test, **the user** is the participant who interacts with the system and verbalises his/her thoughts while doing the tasks.

- There are two major influences that must be taken into account before selecting participants for testing:
  1. Number of participants (sample size)
  2. Relevance of participants.

Number of Participants

- Nielsen (2001) recommends to plan for five participants to find 85% of the problems. But still *controversial*.



Relevance of Participants and Tasks

- **Usability test participants** should be as representative as possible of the targeted users of the tested system.
- Relevant users are more likely to encounter relevant problems, which in turn will produce more relevant results.
- **Usability test tasks** should accurately represent the activities that real users would perform when using an application in order to achieve certain goals.
- Meet representatives from the customer organisation to select the tasks.

Test Observation

- The test evaluators/observers have to observe the user's behaviour and listen to the user's verbalisations in order:

1. Understand the positive and negative aspects of the system, and to detect of usability problems.

2. Record participants' performance measures.

Test Observation Sheet

Usability Test Observation Sheet

Participant #: ____	**TA method:** ____	**Date:** / /2013
Session starts at: ___ h ___ m		**ends at:** ___ h ___ m

Task 1	Task time: ____s	❑ Successful	❑ Unsuccessful
Notes:			
Task 2	Task time: ____s	❑ Successful	❑ Unsuccessful
Preliminary problems discovered			Time problem occurred .....
Task 3	Task time: ____s	❑ Successful	❑ Unsuccessful
Preliminary problems discovered			Time problem occurred .....

- The test sessions can also be videotaped.

Usability Test Environment

- Participants work alone in usability tests, but testing in pairs can be more natural in some situations such as with kids or with systems that used collaboratively.

- Usability tests can be conducted practically anywhere: developments in the areas of computer networks and collaborative work tools mean that even remote testing is possible.

- ***Remote usability testing*** is described as *"usability evaluation where the test evaluators are separated in space and/or time from the test subjects"*.

Usability Test Environment

- In general, however, usability tests are conducted either in specific usability laboratories, or in the field at the customer site.

- The customer site is familiar to the participants, making it easier for them to relax, but is more challenging for evaluators, as interruptions are hard to control, and the available equipment varies from site to site, or has to be brought along specially.

Usability testing

- Specific laboratories, on the other hand, offer dedicated equipment and a peaceful environment, and gives greater control of the variables critically affecting the level of usability, but the participants must then be willing to travel to these laboratories, and the artificial environment can produce unrealistic results.

Pilot Tests

- It is necessary to run a pilot test prior to the actual tests, in order to check the test tasks, instructions, and equipment.

- It should be conducted at most two days before the actual tests are scheduled to take place, so that the preparations are finished but the test team still has enough time to make changes if needed.

Analysing Data and Making Recommendations

- After the test session, the evaluators analyse the data, diagnose the usability problems, and recommend changes to address the problems.

- It is important that evaluators list the problems in order of importance, so that developers can prioritise them accordingly. For example, problems can be classified according to their severity.

- The **severity** of a usability problem refers *to the impact of the problem when it occurs*.

Analysing Data and Making Recommendations

- Dumas and Redish (1999) suggest a four level scale with a clear reference to the impact on users' tasks:

- Level 1 problems: prevent users from completing a task,
- Level 2 problems: significantly slow down the user's performance and frustrate them,
- Level 3 problems: have a minor effect on usability, and
- Level 4 problems: point to potential enhancement in the future.