6001333-3 Human Computer Interaction

Lecture 6 (Week 5)

- Usability

- Usability evaluation

Email: arzahrani@uqu.edu.sa

Defining Usability

- Usability is a key concept in the field of human-computer interaction.

- The term "usability" was coined in 1990s to replace the term "user-friendly".

- The International Standard ISO 9241-11 defines usability as *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use".*

Defining Usability

- "Effectiveness" here to the accuracy and completeness with which users achieve specified goals.

- "Efficiency" means the amount of resources expended in relation to the product's effectiveness.

- "Satisfaction" means that users can complete their tasks without discomfort, and that they feel positive about using the product.
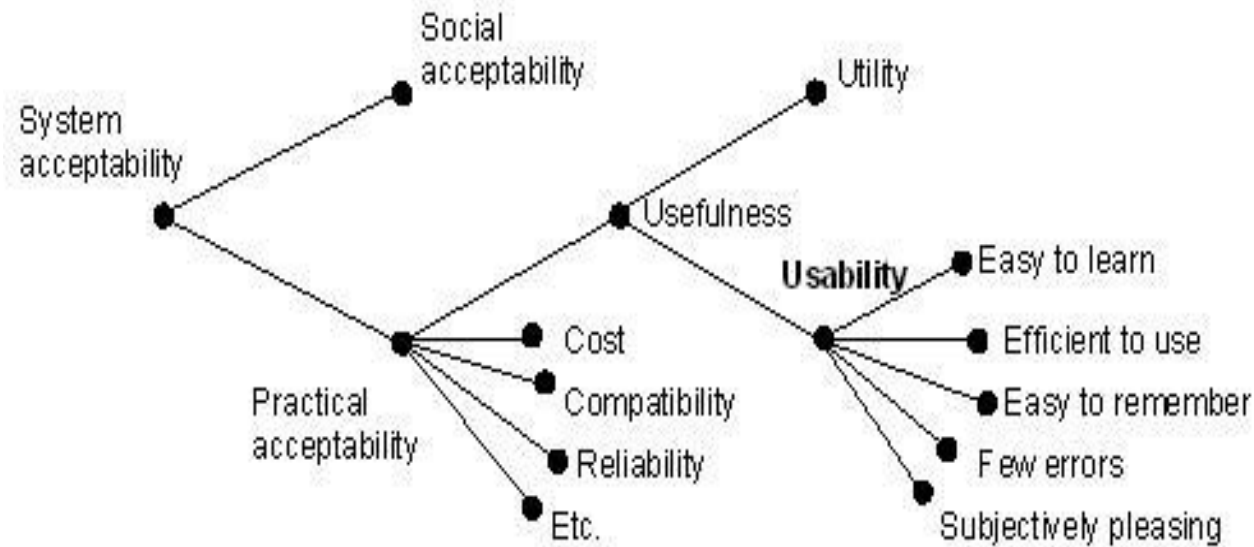
Defining Usability

- Finally, the term "context" includes the users, their goals, the nature of the task(s), and the particular equipment, as well as the physical and social environments in which the product is used.
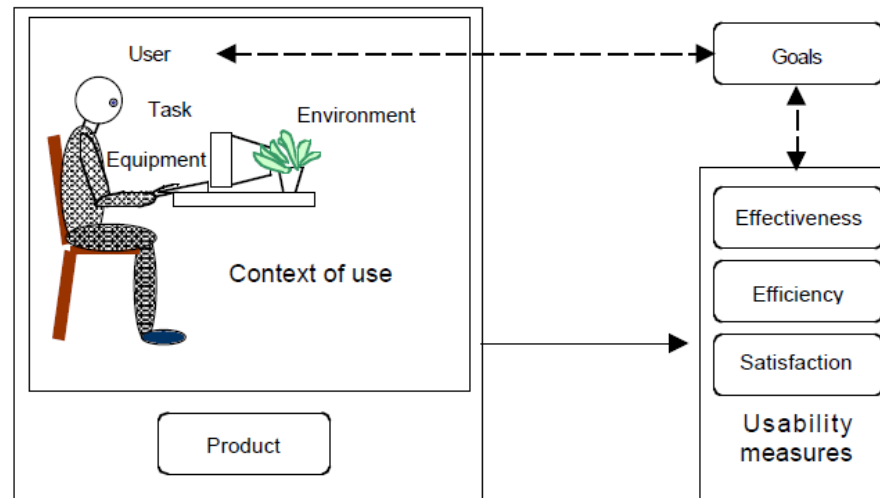
Defining Usability

- Nielsen (1993) defines usability as one of the main elements of a system's acceptability.

Why Usability is Important

- Usability is currently being considered a critical quality factor for software systems, in general, and websites in particular.

- The impact of poor usability on websites can have serious consequences in a competitive environment.

- According to Nielsen (2001), 50% of potential Internet sales are abandoned due to poor website usability. Internet businesses could potentially double their collective sales if e-shops could achieve a better standard regarding the quality of user experience.

Designing Usability

- The International Standard 13407 (1999) provides a framework for designing usable interfaces. This is known as the *usability engineering lifecycle*, and is comprised of four activities:

  1. Understand and specify the context of use;

  2. Specify the user and organisational requirements;

  3. Produce design solutions;

  4. Evaluate designs against requirements

Designing Usability
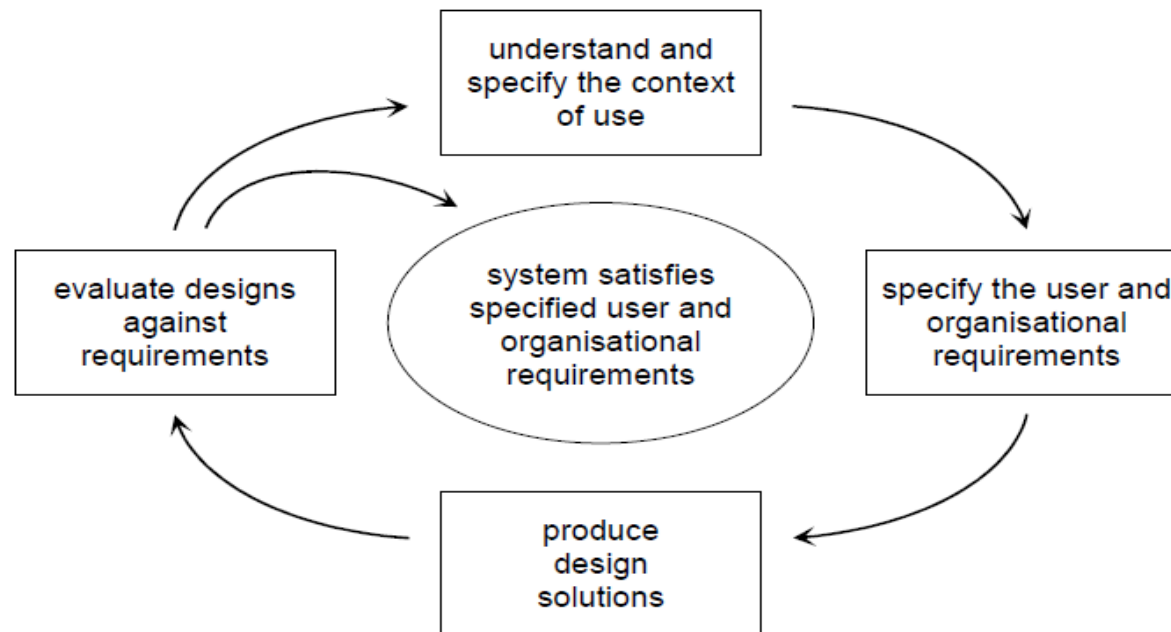
Usability Evaluation

- Koutsabasis et al. (2007) define usability evaluation as 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.

- Usability evaluation is required at several points during the design process. It is, however, important to start evaluation as early as possible, because changes can become more expensive to implement as specific design and functionality decisions are made

Usability Evaluation

- There are two types of evaluation:

1. Formative Evaluation
 – **During** the design and development process to find
   usability problems.

2. Summative Evaluation
  – **After** the design has been implemented to measure its
   effectiveness and efficiency.

Usability Evaluation

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

- **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

- Heuristic evaluation, developed by Nielsen and Molich in 1990, involves inspectors checking whether the interface conforms to a set of guidelines or principles.

- Jakob Nielsen's Ten Usability Heuristics - 10 general principles for user interface design. "

1. Visibility of system status

2. Match between system and the real world

3. User control and freedom

4. Consistency and standards

Heuristic evaluation

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

- It is a cheap, fast, and easy method for evaluating systems.

- A formative method -- it can be used in the early stages of a system's development.

Heuristic evaluation

- **Procedure of the evaluation:**

  The evaluators need to conduct their evaluations separately, and their findings and ratings can then be aggregated after they have all finished the evaluation process. The checklist should be identical for all the evaluators and there should be no differences in terms of either guidelines or instructions.

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

Heuristic evaluation

- The "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).

- In Jacobsen et al. (1998) study, four evaluators were asked to list and describe all the usability problems that they could detect from an interface.

Heuristic evaluation

- The results showed that each evaluator detected between 39% and 63% of the total number of problems; only 20% of the problems were detected by all the evaluators, and 46% were detected by only one evaluator.

- The any-two agreement formula provided by Hertzum and Jacobsen (2001) can be used to calculate inter-coder reliability across.

Heuristic evaluation

$$Any - two\ agreement = \frac{|Pi \cap Pj|}{|Pi \cup Pj|}$$

- In this equation, Pi and Pj are the problems identified by evaluators "i" and "j" respectively. Its value ranges from 0% in the case of no agreement amongst the evaluators to 100% in the case of full agreement.

Cognitive Walkthrough

- 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.

Cognitive Walkthrough

- To do a walkthrough you need four things:

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.

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?

Model-based Methods

- Model-based methods in usability evaluation are the least commonly used of the three methods.

- They stem from psychological research into human performance.

- The primary aim of adopting these methods is to predict certain aspects of user performance with an interface, such as total task time, or the difficulty of learning a task's sequence.

GOMS

- A good example of a model-based method is the GOMS (Goals, Operators, Methods and Selection Rules) model, which can be used to estimate the amount of time required for completing certain tasks.

- Probably the most widely known and used technique in this family.

- Developed by Card, Moran and Newell in 1983.

GOMS

- **<u>Goals</u>**:

  - End state trying to achieve.

  - Then decompose into subgoals.
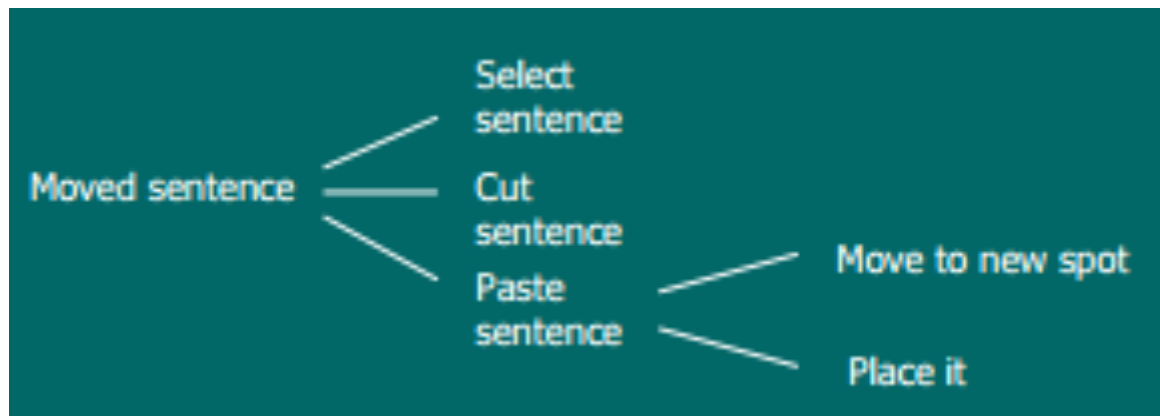
GOMS

- **<u>Operators</u>**:

  - Basic actions available for performing a task (lowest level actions).

  - Examples: move mouse pointer, drag, press key, read, dialog box, …

GOMS

- **<u>Methods</u>**:

- Sequence of operators (procedures) for accomplishing a goal

Example: Select sentence

 – Move mouse pointer to first word

 – Depress button

 – Drag to last word

 – Release

GOMS

- **Selection rules:**

  - Invoked when there is a choice of a method

  - GOMS attempts to predict which methods will be used

  - Example: Could cut sentence either by pulldown or by ctrl-x

  - GOMS Assumes error -free performance.

GOMS

Move Sentence

1. Select sentence
Reach for mouse	H	0.40
Point to first word	P	1.10
Click button down	K	0.60
Drag to last word	P	1.20
Release	K	0.60
		3.90 secs

User-based Methods

- Many methods exist for conducting user-based evaluation, such as surveys, interviews, and focus groups. Another approach is to conduct usability testing using behavioural observation, such as think-aloud protocol.

- 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. As a result, behavioural observation such as think-aloud protocol is more widely used.

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. The think-aloud approach has been developed in response to this challenge.

Usability testing

- The general idea is for test participants to 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 TA method is that it can be time-consuming and expensive compared to expert-based or model-based evaluation methods.

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.

Usability testing

- 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, whereas preference measures indicate how much the users enjoy using the system.

Usability testing

- Interestingly, a number of studies have found low correlations between user performance and user preference measures.

- In a TA 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: number of participants (sample size), and relevance of participants.

Usability testing

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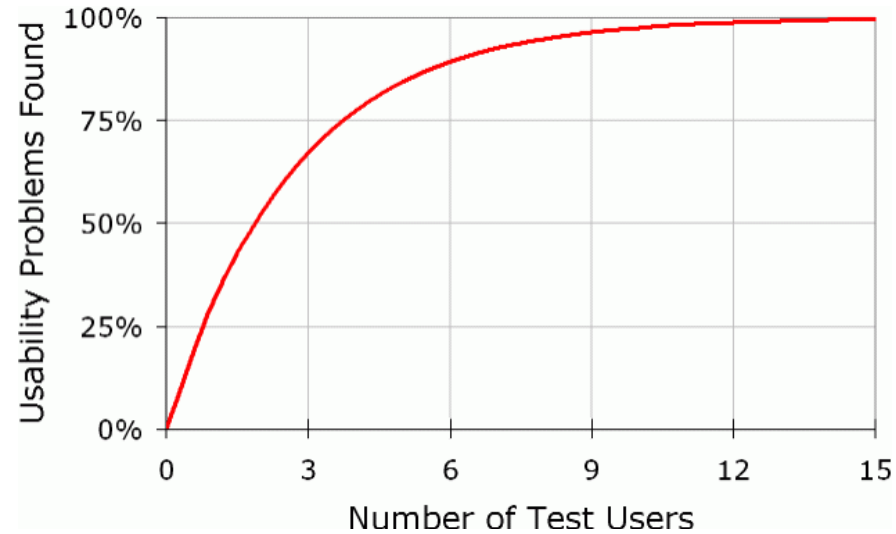
Usability testing

- There is general agreement between usability researchers that, regardless of size, a test sample 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 testing tasks** should accurately represent the activities that real users would perform when using an application in order to achieve certain goals.

Usability testing

- Group with representatives from the customer organisation to select the tasks.

- Apart from allocating tasks and giving instructions, the evaluator also needs to "read the user". This means that he/she has to observe the user's behaviour and listen to the user's verbalisations in order to understand the positive and negative aspects of the system, and to achieve the goal of usability testing the detection of usability problems.

Usability testing

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 testing

- 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.

Usability testing

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

- Remote usability evaluation can be separated into two key categories; synchronous and asynchronous methods.

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

Usability testing

- The real use context, with tasks emerging from the users' work, reveals problems that would be hard to detect in laboratory settings with predefined tasks.

- 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.

- Specific laboratories, on the other hand, offer dedicated equipment and a peaceful environment, but the participants must then be willing to travel to these laboratories.

Usability testing

- In addition, the artificial environment can produce unrealistic results. Nonetheless, testing in laboratories gives greater control of the variables critically affecting the level of usability, and the measurements obtained are more precise than in the field tests

Usability testing

- It is necessary to run a pilot test prior to the actual tests, in order to check the test tasks, instructions, and equipment. The pilot participant does not have to be from the target group, but should be somebody who is not part of the evaluation team.

- Dumas and Redish recommend that the pilot test is conducted 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.

Usability testing

- 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.

Usability testing

- 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.

Usability testing

- 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.