Concepts of Usability Engineering

2022

Prof Judy Kay | Changhao Li *

Content

*This pamphlet's content is adapted from COMP5427: Usability Engineering at the University of Sydney. The course instructor is Prof Judy Kay in Semester 2, 2022, and the author Changhao Li merely extracted the essence of her course content. If you found typos and other errors, feel free to reach Changhao Li via changhaoli.gz@outlook.com for content corrections. Have fun in HCI!

1 Introduction

Definition 1.1: Usability

Usability includes:

Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design?

Efficiency: Once users have learned the design, how quickly can they perform tasks?

Memorability: When users reuse product after a long time, how easily can they reestablish proficiency?

Errors: How many and how severe are these errors, and how easily can they recover from the errors?

Satisfaction: How pleasant is it to use the design?

Definition 1.2: Utility

Utility indicates whether the product provides the features the user needs.

Argument 1.1: Usefulness

Usefulness = Usability + Utility

Definition 1.3: User Experience

User Experience means more than Usefulness (which only focus on performance), including...

- Emotional and visceral response
- Fun (hedonic response), exciting
- Artistic appeal
- Match to fashions and user expectations (or not): surprising, boring
- Gestalt

2 Terminology

Definition 2.1: Empirical Study

In usability engineering, *empirical study* means that these are studies that involve people. You observe them, ask them questions and ask them to do questionnaires.

Definition 2.2: WEIRD

Participants of empirical studies are often drawn from Western, Educated, Industrialized, Rich, Democratic populations, causing selection bias.

Definition 2.3: Critical incident

A critical incident is the single most important source of qualitative data in formative evaluation.

Argument 2.1: Science v Engineering v Design

Science: Discover things that have long existed ("is").
Engineering: Create things that did not previously exist ("will be").
Design: Art.

Argument 2.2: Formative v Summative

Formative evaluation: a diagnostic evaluation to improve user interface on the next iteration of the cycle.
Summative evaluation: used to make a judgement.

Argument 2.3: Slips v Mistakes

Slips: correct user model, inadvertent incorrect action.
Mistakes: incorrect mental model.

3 Mental Model

Argument 3.1: "Designers cannot think like Users"

Designers may need to understand the domain, the context of use, what the user knows, what they have experienced, how they will interpret the interface elements and what they will see.

Other reasons might include challenges of creating specifications; some interfaces need to be complex; some are complex –even though most people do not need all the functionality; tension between utility and beauty, fun, and marketing; challenges of making interfaces fit for different purposes and people in different contexts or cultures with different devices; evolution of purposes and of software; reuse may be limited by IP (legal issues), etc.

Definition 3.1: Mental Model

A mental model is what people believe.

Mental models define what people see, drive predictions, and plan actions.

Warning:

How to address mental model problems?

- Fix the system - make it match users' mental models.
- Fix the user - improve users' mental models so that they more accurately reflect your system.

4 User Goal/Task

Definition 4.1: User Goal

User Goal is user requirements.

Definition 4.2: User Task

User Task, aka Benchmark Test, is concrete things users want to do.

Argument 4.1: Good User Tasks!!!

Each user task should have these properties:

- They are **concrete**

  User understands what they need to do:

  instructions are clear,

  easily understood,

  provide all the information with the user needs,

  different people should interpret them in the same way.

- You can **judge success**

  Clear start and clear end when you can assess users' success.

- They **do not lead** the user

  Tell WHAT to do, not HOW to do the task. Avoid any words that appear at the interface.

- They are **relevant** to user goals for the interface

  Users would expect to be able to do them to achieve user goals.

The set of tasks should have these properties:

- Together they give **good coverage**

  Frugal: each task tests different things (unless you explicitly want to study repeat use)

  Effective: each task covers an important aspect

- They are at the right level of **difficulty**

  Start with easy tasks

- They are **respectful** and avoid offending the user

  Humour is dangerous

5 Contextual Inquiry

Definition 5.1: Auto-ethnography

Auto-ethnography means observing oneself in authentic settings.

The purpose of auto-ethnography is that it gives insight into user experience of learning about something. But be aware: other people's experiences would be different from yours.

Definition 5.2: Contextual Inquiry (CI)

Contextual Inquiry is a user-centred research method related to auto-ethnography. Researchers work in the real scene of working product and understand the workflow of users via **observation** and **interview**.

Steps of Contextual Inquiry:

1. Identify stakeholders
2. Identify problems they face that the new system should help overcome
3. Identify tasks people should be able to do with the new system

For each group of stakeholders:

- Observe them in action
- Interview them
- Use this to create "work activity notes"

CI changes user goals and tasks accordingly. It helps form the product's System Concept Statement and Hierarchical Task Analysis model。

6 Persona

Definition 6.1: Persona

A persona represents a major user group for your system. It expresses and focuses on the major needs and expectations of the most important user groups.

Persona describes real people with backgrounds, goals, and values.

Argument 6.1: Elements of Persona

- Persona group
- Fictional name
- Job titles and major responsibilities
- Demographics such as age, education, ethnicity, and family status
- The user goals they are trying to complete using the system
- Their physical, social, and technological environment
- A quote that sums up what matters most to the persona
- Casual pictures

Argument 6.2: Building Persona

Observe:

Study users to understand why they need it and what they need it to do for them

Ask:

Interviews, questionnaires

Experience:

Try doing what your intended users do

Literature

Use the work of others who have done the above steps

Tips:

Persona is useful because:

It helps in discussions in the design group and making design trade-offs;

Designs can be constantly evaluated against the personas.

Definition 6.2: Trans-Theoretic Model (TTM) of Behaviour Change

Precontemplation, Contemplation, Preparation, Action, Maintenance.

7 Prototype

Definition 7.1: Prototype

A prototype is a draft version of a product that allows you to explore your ideas and show the intention behind a feature or the overall design concept to users before investing time and money into development.

From lo-fi to hi-fi:

- Paper-based sketches
- Paper-based storyboard
- Computer-aided sketches/storyboard
- Wizard of Oz / Video-prototyping
- Computer-based scenario simulation
- Computer-based Horizontal simulation
- Computer-based Vertical simulation
- Computer-based full functionality simulation

8 Think Aloud

Argument 8.1: Think Aloud is THE usability tool

Think Aloud is THE usability tool. (Jakob Neilsen)

Definition 8.1: Think Aloud

Users do benchmark tests and say what they are thinking at the same time. Designers take notes and revise the product interfaces accordingly.

Think aloud is practised many times in classes and assignments.

9 Heuristic Evaluation

Definition 9.1: Heuristic Evaluation (HE)

Heuristic Evaluation is a process where experts use rules of thumb (e.g. Neilsen's heuristics) to measure the usability of user interfaces in independent walkthroughs and report issues.

HE establishes a set of user tasks, selects a set of design heuristics, and recruits HE experts. Each expert needs to walk through tasks, meanwhile, team member records violations of the heuristics. Expert reviews all the problems found and rate them.

Argument 9.1: Factors of Severity Rating

- The frequency with which the problem occurs: Is it common or rare?
- The impact of the problem if it occurs: Will it is easy or difficult for the users to overcome?
- The persistence of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem?"

The results of HE can be represented by severity rating. The rating of severity rating range from 0 to 4, where 0 means the interface has no problem whereas 4 means that the interface has severe problems.

Tips:

Neilsen's 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.

If there are less than 5 participants, then heuristic evaluation can be called *discount evaluation*.

10 Questionnaire

There are many types of questionnaires, such as post-task v post-study, usability v hedonic, closed v open, etc.

Argument 10.1: Asking v Observing

A questionnaire is based on asking, it's subjective;
Others such as Think Aloud are based on observing, they are objective.

Warning:

Why is it important to use standard questionnaires in HCI?
- Reliability: consistent responses to the questions. (TEST: ask the same question more than once)
- Validity: correlates with other evidence, e.g. completion rates.
- Sensitivity: discriminates between good and bad interfaces. #<10 ... SUS/SEQ
- Objectivity: give an independent instrument.
- Quantification: fine grain of reporting and statistical analysis.
- Economy: cheap to design.
- Comparability: scientific generalisation

Definition 10.1: System Usability Scale (SUS)

The participant answers the following questions:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome (awkward) to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

SUS measures a product's learnability and usability. Usually, it needs 8-12 participants.

Definition 10.2: UMUX

The participant answers the following questions:

1. This system's capabilities meet my requirements.
2. Using this system is a frustrating experience.
3. This system is easy to use.
4. I have to spend too much time correcting things with this system.

Definition 10.3: UMUX-lite

The participant answers the following questions selected from UMUX:

1. This system's capabilities meet my requirements.
2. This system is easy to use.

Definition 10.4: Net Promoter Score (NPS)

The participant answers the following question:

- How likely are you to recommend this product to a friend?

Definition 10.5: Single Ease Question (SEQ)

The participant answers the following single question:

- Overall, this task was?

Definition 10.6: Simple Open Questionnaire (SOQ)

The participant answers the following open questions:

1. What was the best thing about this interface?
2. If you could change one thing in this interface what would it be?
3. Any other comments?

11 Cognitive Walkthrough

Cognitive Walkthroughs (CW) access learnability part of usability.
More specifically, CW assesses user success and recovery from errors.

Definition 11.1: Cognitive Walkthrough (CW)

User has task

0. Will user understand this sub-task is needed?

Explores system for useful elements

1. Will correct action be obvious?
2. Will user understand instructions?

Selects one to try

User interprets system response

3. Will user interpret machine action correctly?

Benefits of CW includes its low cost and no users needed.

Definition 11.2: Compact Cognitive Walkthrough (CCW)

Ask ourselves (expert perspective) when doing tasks:

1. Would our persona know what to do at this step?
2. If the persona does the right action, do they know they have made progress toward their goal?

Definition 11.3: Simplified Streamlined Cognitive Walkthrough (SSCW)

SSCW reduces training time and increases the researcher's responsibility instead of the designer's.

Definition 11.4: Simplified Pluralistic Walkthrough (SPW)

SPW mostly listened and asked clarifying questions to the participants.

Warning:

SPW is similar to Think Aloud because both involve users and both require a well-designed set of tasks (as do in HE, CW, and GOMS).

12 HCI Laws

Definition 12.1: Fitt's Law

The formula representing Fitt's Law is

$$MT = a + b\log_2(\frac{D}{W} + 1)$$

...where D is distance, W is width. Therefore, the time to move quickly to a target:

- Further takes longer
- Smaller targets take longer

Note that *Magic Pixel* is Prime Pixel (cursor's location right now) plus 4 corners on the screen.

Definition 12.2: Tesler's "Law"

Tesler argues that engineers should put more time into simplifying an application instead of making millions of users spend an extra minute using the more complex system.

Tesler's argument is also called conservation of complexity.

Definition 12.3: Miller's Law (Magic 7)

Most people can only retain 7 ± 2 "chunks" of information in short term memory at any given time.

More specifically, 7 for numbers, 6 for letters and 5 for words.

Definition 12.4: Hick's "Law"

Hick observed that the time to choose between equally likely alternatives increases as the log of the number of choices:

$$t = b\log_2(n + 1)$$

Hick's observation does not apply when alternatives are not equally likely.

Definition 12.5: Power of Practice Law

Time to do many mechanical and cognitive skills gets faster with repetition.

$$T_n = T_1 \times n^{-a}$$

Practice makes perfect.

13 GOMS

GOMS measures the speed performance of experts conducting user tasks.

Definition 13.1: GOMS

GOMS is a model representing the interaction knowledge and cognitive process of the user and the interface. Here G is Goal, O is Operation, M is Method and S is the Selection rule.

GOMS is relevant to the interfaces for the small core of tools that people use extensively or need automation, such as teacher's marking or input text online, etc.

GOMS's role is to design the interaction methods and to decide what interaction methods to implement.

GOMS requires highly specialised expert evaluators, does not involve users and it assumes users are experts. GOMS is very tedious since it's impractical to do it by hand. It uses professional automated tools.

Tips:

GOMS mainly evaluates 1st (Command Line) and 2nd (WIMP) generation of the user interface.

Definition 13.2: WIMP

Windows, Icons, Menus, Pointer.

Keystroke analysis is related to GOMS.

Definition 13.3: Keystroke Analysis

Keystroke Analysis' steps are: 1) Given a task, define the methods to be considered; 2) For each method, define the sub-tasks; 3) For each sub-task, do the keystroke analysis.

The components of Keystroke Analysis include:

K = key press [0.08-1.20s]

P = point with mouse [0.80-1.50s (Fitts Law)]

C = click with mouse [0.20s]

H = switch device [0.40s]

M = mentally prepare [1.35s]

R(t) = system response time [?s]

During the final exam, choose mediocre values instead of extreme values.

14 Accessibility

Definition 14.1: Accessibility

Web accessibility means enabling web access to people with disabilities.

Argument 14.1: POUR Principle

Perceivable: Can users perceive content?

Operable: Can users use UI components and navigate the content?

Understandable: the core of usability generally...

Robust: Can it work with diverse user agents (browsers)?

The POUR Principle came from Web Content Accessibility Guidelines (WCAG 2.0).

Definition 14.2: Alternative Texts

For people who rely on screen readers because they cannot see images online, the image needs a description, aka alternative text, aka alt-text.

15 A/B Testing

A/B Testing is an example of an empirical form of field study beyond the scope of this unit.

Definition 15.1: A/B Testing

A/B Testing compares subtly different versions of interfaces by controlled experiment.

The process of A/B Testing can be concluded as:

1 Driving question

2 Hypothesis

3 Calculate time to run test

4 Run the test

5 Review results