Usability Evaluation Methods II

SPRING 2020

DR MARK FREEMAN

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Week	Topic	Reading
1	Introduction to HCI; Design Principles	Chapter 1
2	User-Centred Design Process	Chapter 2 & 3 / Gould et al. (1987)
3	User Interaction and Interfaces	Chapters 4, 5 & 6 / Shin et al. (2017)
4	Interaction Design and Development I	Chapters 7 & 8
5	Interaction Design and Development II	Chapters 9 & 10
6	Interaction Design and Development III	Chapters 11 & 12
7	Information Presentation and Design Patterns	
8	Usability Evaluation Methods I	Chapters 13 / Borsci et al. (2015)
9	Usability Evaluation Methods II	Chapter 14, 15 & 16
10	Accessibility and Special Issues in HCI	Online: WCAG2.0
11	Models, Theories and Risks	MacKenzie (1992)
12	Mixed Reality and Future HCI	
13	Subject Revision	

This Week

Usability Evaluation Methods

Expert-Based Methods

- Heuristic Evaluations
- Walkthroughs

Predictive Models

Subject Description

- The subject provides students with an understanding of Human Computer Interaction (HCI) principles and practices, and how to apply them in the context of developing usable interactive computer applications and systems. The subject also emphasises the importance of taking into account contextual, organisational, and social factors in the design of computer systems. Students will be taken through the analysis, design, development, and evaluation of user interfaces. They will acquire hands-on design skills through an interaction design project. The subject will cover topics including user-centred design, the development process, prototyping, usability testing, measuring and evaluating the user experience and accessibility.

Subject Learning Outcomes (SLOs)

- On successful completion of this subject, students will be able to:
    1. Identify and describe HCI principles and design issues.
    2. Discuss and justify HCI solutions based on design principles.
    3. Demonstrate an understanding of the HCI design process.
    4. Acquire skills to design and implement user-centred design.
    5. Select and use suitable methods of measuring and evaluating the user experience.

Remember: After a system has been developed…

THIS WEEK WE NEED TO SHIFT OUR THINKING

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Usability Evaluation Methods (UEMs)

PREECE ET AL. (2016)

Quick and Dirty Evaluations

- Informal (throughout UCD)

Usability Testing

- Measuring user performance

Field Studies

- In natural settings

Predictive Evaluation

- Expertise

Usability Evaluation Methods (UEMs)

- User-Based Evaluation
  - Performance, non-verbal behaviour, attitude, cognition, stress and motivation
  - user experience and the interconnectivity that exists between usability and user experience (Vermeeren et al., 2010)
    - *(Quick and Dirty, Usability Testing, Field Studies)*
- Expert-Based Evaluation
  - Conformance and attitude
    - *(Quick and Dirty, Predictive Evaluation)*
- Theory Based Evaluation
  - Performance (idealised)
    - *(Predictive Evaluation)*

(Sweeney et al. 1993)

Evaluation methods

Method	Controlled settings	Natural settings	Without users
Observing	X	X	
Asking users	X	X	
Asking experts		X	X
Testing	X	x*	
Modeling			X

* Some testing can be performed but with greater difficulty

Expert Inspections

Experts use their knowledge of users & technology to review software usability.

Expert critiques can be formal or informal.

Heuristic evaluation is a review guided by a set of heuristics.

Walkthroughs involve stepping through a pre-planned scenario noting potential problems.

Traditional Heuristic Evaluation

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Traditional Heuristic Evaluation

- Developed by Jacob Nielsen in the early 1990s.
- Based on heuristics distilled from an empirical analysis of 249 usability problems.

- These heuristics have been revised for current technology by Nielsen and others for:
  - mobile devices,
  - wearables,
  - virtual worlds, etc.

- Design guidelines form a basis for developing heuristics.

Jakob Nielsen’s

TEN USABILITY CHARACTERISTICS

- Visibility of system status
  - The system should always keep users informed about what is going on, through appropriate feedback 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. 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 dialogue. Support undo and redo.
- Consistency and standards
  - Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

https://www.nngroup.com/articles/ten-usability-heuristics/

Jakob Nielsen's
TEN USABILITY CHARACTERISTICS

- Error prevention
  - Even better than good error messages is a careful design which 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 dialogue 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.

https://www.nngroup.com/articles/ten-usability-heuristics/

Jakob Nielsen's
TEN USABILITY CHARACTERISTICS

- Aesthetic and minimalist design
  - Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue 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 (no 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.

https://www.nngroup.com/articles/ten-usability-heuristics/

Revised Version (2014) of Nielsen's Original Heuristics

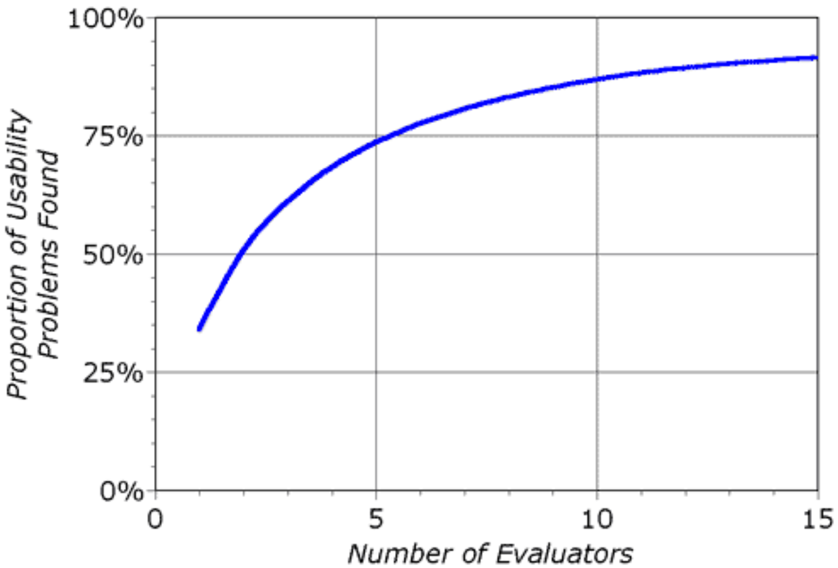
Tool

- Visibility of system status
- Match between system and real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognize, diagnose, recover from errors.
- Help and documentation

Number of Evaluators

- Nielsen suggests that on average 5 evaluators identify 75-80% of usability problems.
- Cockton and Woolrych (2001) point out that the number of users needed to find 75-80% of usability problems depends on the context and nature of the problems.

Number of Evaluators in a Heuristic Evaluation



https://media.nngroup.com/media/editor/2012/10/30/heur_eval_finding_curve.gif

Advantages of Heuristic Evaluation

- Evaluators can focus their attention on certain issues as identified by the developers and /or project managers
- Heuristic evaluation does not carry the ethical and practical issues/problems associated with inspection methods involving real users
- Evaluating designs using a set of heuristics can help identify usability problems with individual elements and how they impact the overall user experience

Issues with Heuristic Evaluations

- Can be difficult & expensive to find experts
- Choosing appropriate heuristics is important
- Time-consuming when compared to other 'quick and dirty' inspection methods
- Based on preconceived ideas of what makes a system usable
- Best experts need to have knowledge of both the application domain & users
- Neilsen's Heuristics are abstract

- Biggest problems:
  - Important problems may get missed
  - Many trivial problems are often identified (false alarms)
  - Experts have biases

3 Stages for Conducting a Heuristic Evaluation

1. Briefing session to tell experts what to do.
2. Evaluation period of 1-2 hours in which:
   - Each expert works separately
   - Take one pass to get a feel for the product
   - Take a second pass to focus on specific features

3. Debriefing session in which experts work together to prioritize problems.

Shneiderman’s 8 Golden Rules of Interface Design

- Strive for consistency
- Enable frequent users to use shortcuts
- Offer informative feedback
- Design dialogue to yield closure
- Offer simple error handling
- Permit easy reversal of actions
- Support internal locus of control
- Reduce short-term memory load

- Students to read: Wong (2017) Shneiderman’s Eight Golden Rules Will Help You Design Better Interfaces
  - https://www.interaction-design.org/literature/article/shneiderman-s-eight-golden-rules-will-help-you-design-better-interfaces

Heuristics for Websites Focus on Key Criteria (Budd, 2007)

- Clarity
- Minimize unnecessary complexity & cognitive load
- Provide users with context
- Promote positive & pleasurable user experience

Modern Heuristic Evaluations

Tool

- Nielsen’s and Shneiderman’s lists can be used as a starting point by experts.

- Design specific heuristics should be created, consider:
  - User
  - Task (goal)
  - Environment

Interaction Design Foundation – Method

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**
5. **Second evaluation phase**
6. **Record problems**
7. **Debriefing session**

(https://www.interaction-design.org/literature/article/heuristic-evaluation-how-to-conduct-a-heuristic-evaluation)

Tool

Cognitive Walkthroughs

"THEY ARE DESIGNED TO SEE WHETHER OR NOT A NEW USER CAN EASILY CARRY OUT TASKS WITHIN A GIVEN SYSTEM. IT IS A TASK-SPECIFIC APPROACH TO USABILITY (IN CONTRAST TO HEURISTIC EVALUATION WHICH IS A MORE HOLISTIC USABILITY INSPECTION). THE IDEA IS THAT IF GIVEN A CHOICE – MOST USERS PREFER TO DO THINGS TO LEARN A PRODUCT RATHER THAN TO READ A MANUAL OR FOLLOW A SET OF INSTRUCTIONS." (IDF, 2016)

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Cognitive Walkthroughs

- Focus on ease of learning
- Designer presents an aspect of the design & usage scenarios
- Expert is told the assumptions about user population, context of use, task details
- One or more experts walk through the design prototype with the scenario
- Experts are guided by 3 questions:
  1. **Will the correct action be sufficiently evident to the user?**
  2. **Will the user notice that the correct action is available?**
  3. **Will the user associate and interpret the response from the action correctly?**

  As the experts work through the scenario they note problems.

Cognitive Walkthrough for the Web (CWW), Blackmon et al. (2002)

- CWW uses contextually rich descriptions of user goals (100-200 words long) incorporating more information about users’ understanding of their tasks and underlying motivation
- CWW assumes that generating an action (e.g., clicking on a link, button, or other widget) is a two step process.
  - Step 1: involves parsing a new page into subregions and attending to the correct subregion of the page
  - Step 2: selecting a widget from the attended
- CWW evaluation process is organized differently and fits better into website development
- A user of CWW works on one web page at a time in relation to a whole set of representative user goals

Cognitive Walkthrough for the Web (CWW) Blackmon et al. (2002)

- Changes to the 3 questions
  - Q2a) Will the user connect the correct subregion of the page with the goal using heading information and her understanding of the sites page layout conventions? and
  - Q2b) Will the user connect the goal with the correct *widget* in the attended to subregion of the page using link labels and other kinds of descriptive information?

Pluralistic Walkthrough

- Variation on the cognitive walkthrough theme
- Performed by a carefully managed team
- The panel of experts begins by working separately
- Then there is managed discussion that leads to agreed decisions
- The approach lends itself well to participatory design

Pluralistic Walkthrough – Different View

- “A usability test method employed to generate early design evaluation by assigning a group of users a series of paper-based tasks that represent the proposed product interface and including participation from developers of that interface.”
- “A systematic group evaluation of a design in which usability practitioners serving as walkthrough administrators guide users through tasks simulated on hard-copy panels and facilitate feedback about those tasks while developers and other members of the product team address concerns or questions about the interface.”
-

http://www.usabilitybok.org/pluralistic-walkthrough

Predictive Models

PREDICTIVE MODELS EVALUATE A SYSTEM WITHOUT USERS BEING PRESENT

PREDICTIVE MODELS USE FORMULAS TO DERIVE VARIOUS MEASURES OF USER PERFORMANCE

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Predictive models

- Provide a way of evaluating products or designs without directly involving users
- Less expensive than user testing
- Usefulness limited to systems with predictable tasks - e.g., telephone answering systems
- Based on expert error-free behavior

GOMS Model

- Developed in the 1980s by Card, Moran and Newell
- A way to model the knowledge and cognitive processes when a user interacts with a system
- GOMS
  - **Goals** - particular state the user wants to achieve
  - **Operators** - the cognitive processes and physical actions that need to be performed to attain those goals
  - **Methods** - learned procedures for accomplishing the goals (the exact steps required)
  - **Selection rules** - used to determine which method to select when there is more than one available for a given stage of a task

GOMS Versions

- CMN-GOMS – Card, Moran, and Newell (1983)
  - Designed to express a goal and sub-goals in a hierarchy, methods and operators and how to formulate selection rules
- KLM – Keystroke Level Model
  - simplified version CMN, focuses on keystrokes and mouse movements + simple mental operations (no goals, methods or selection rules)
- NGOMSL – (1986)
  - a procedure for identifying all the GOMS components, expressed in a form similar to an ordinary computer programming language.
- CPM-GOMS – (1990)
- uses **c**ognitive, **p**erceptual, and **m**otor operators in a **c**ritical-**p**ath **m**ethod schedule chart (PERT chart) to show how activities can be performed in parallel.

John, B. E., & Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: Which technique?. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *3*(4), 287-319.

Keystroke Level Model (KLM)

- Quantitative modelling tool
- Designed to predict in an idealised situation how long a skilled would take to complete a task with no errors

- Times set for physical and mental operators, e.g.:
  - Keystrokes
  - Button clicks
  - Mouse movements
  - Keyboard to mouse movements
  - Mental operations (thinking)

Prior Historical Research

- **Homing**: Moving Hand to Keyboard or Mouse: 360ms
- **Clicking:** the Mouse: 230 ms
- **Pointing:** with the Mouse: 1100ms
- **Mental Operations:** (Deciding what to Do): 1350 ms

- Concept using KLM prediction for a skilled user within 10-20% of the actual time

From Text

Operator name	Description	Time (s)
K	Pressing a single key or button Skilled typist (55 wpm) Average typist (40 wpm) User unfamiliar with the keyboard Pressing shift or control key	0.35 (average) 0.22 0.28 1.20 0.08
P	Pointing with a mouse or other device to a target on a display	1.10
P_1	Clicking the mouse or similar device	0.20
H	Homing hands on the keyboard or other device	0.40
D	Draw a line using a mouse	Variable depending on the length of line
M	Mentally prepare to do something, e.g. make a decision	1.35
R(t)	System response time – counted only if it causes the user to wait when carrying out his/her task	*t*

$$T_{execute} = T_K + T_P + T_H + T_D + T_M + T_R$$

Example from Text

- Change
  - *Running through the streets at night is normal.*
- *To*
  - *Running through the streets at night is not normal.*

The times for each of these operators can then be worked out:

Mentally prepare (M)	1.35
Reach for the mouse (H)	0.40
Position mouse before the word 'normal' (P)	1.10
Click mouse (P_1)	0.20
Move hands to home position on keys (H)	0.40
Mentally prepare (M)	1.35
Type 'n' (good typist) (K)	0.22
Type 'o' (K)	0.22
Type 't' (K)	0.22
Type 'space' (K)	0.22
Total predicted time:	5.68 seconds

When there are many components to add up, it is often easier to put together all the same kinds of operator. For example, the above can be rewritten as

$$2(M) + 2(H) + 1(P) + 1(P_1) + 4(K) = 2.70 + 0.80 + 1.10 + 0.2 + 0.88 = 5.68 \text{ seconds.}$$

Examples from MeasuringU

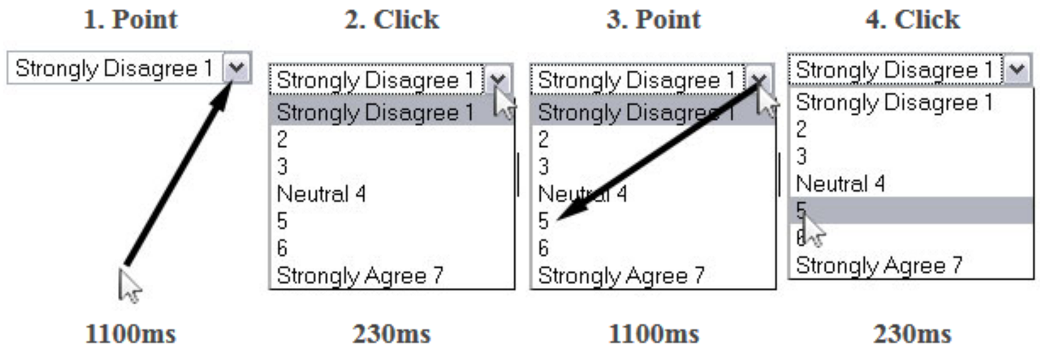


Figure 1: Selecting an option from a drop-down list will take an experienced user 2660ms (2.66 seconds).

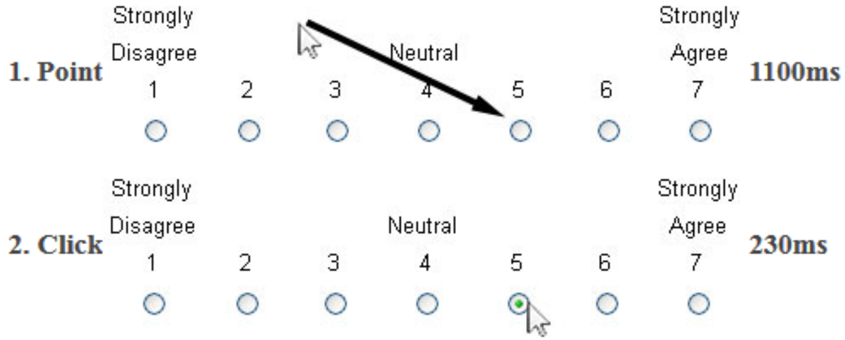


Figure 2: Selecting an option from a radio-button group will take an experienced user 1330ms (1.33 seconds).

https://measuringu.com/predicted-times/

Fitt’s Law

- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, *47*(6), 381.

- Predictive Model of human movement
- Designed to predict the time required to move to a target area is a function of the ratio of the distance to the target and the width of the target

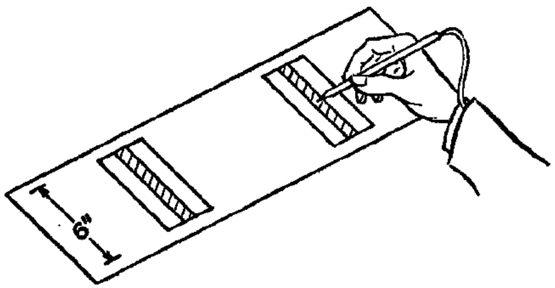


FIG. 1. Reciprocal tapping apparatus. The task was to hit the center plate in each group alternately without touching either side (error) plate.

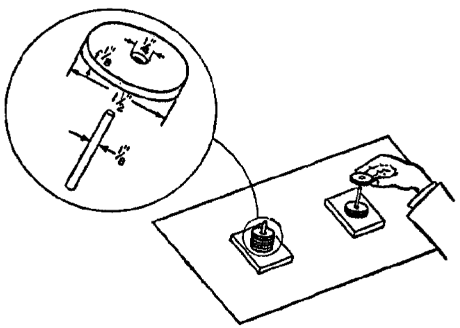


FIG. 2. Disc transfer apparatus. The task was to transfer eight washers one at a time from the right to the left pin. The inset gives the dimensions for the $W_s = \frac{1}{8}$ in. condition.

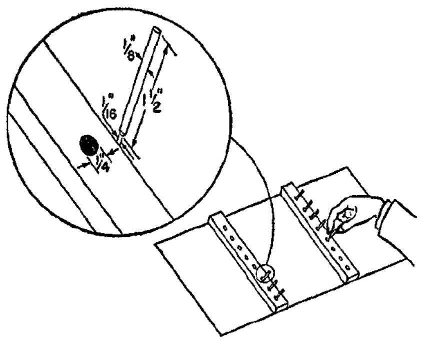


FIG. 3. Pin transfer apparatus. The task was to transfer eight pins one at a time from one set of holes to the other. The inset gives the dimensions of pins and holes for the $W_s = \frac{1}{8}$ in. condition.

Fitt’s Law from Lynda.com

https://www.youtube.com/watch?v=95RoKSFyQ_k

Judge

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Benefits of GOMS

- Allows for comparative analysis of different interfaces
- Can be adapted for new and innovative products
    - E.g. Marco Attention Shift (SMarco) – time to shift attention from a mobile device to a distant object in the real world (Holleis et al. 2007)

Limitations of GOMS

- Narrow focus on efficiency and time
- Level of practicality
- No user involvement
- Only works for routine data-entry tasks
- Predicts idealised performance
- Does not model errors
- Does not consider the user’s state (e.g. fatigue, multitasking)

Main Limitation:

Predictive models can only make predictions about predictable behaviours

Thus, difficult to evaluate how systems will work in the real-world

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

What is your favourite UEM? Why?

Judge

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Take Home Message…

What should you test/when and with who?

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Questions

UNIVERSITY
OF WOLLONGONG
AUSTRALIA