



# Design Process: Evaluating Interactive Systems



# What is Evaluation

- Evaluation is defined as .. “To examine and judge carefully”
- In order to examine and judge, we need criteria against which to base our assessment.



---



# Why Evaluate



- In HCI we evaluate interfaces and systems to:
    - Determine how usable they are for different user groups
    - Identify good and bad features to inform future design
    - Compare design choices to assist us in making decisions
    - Observe the effects of specific interfaces on users
-

---



## Criteria for Evaluation

- Expert Analysis
- User-Based Criteria
- Model Based





# Expert Analysis

**Heuristic Evaluation**  
**Cognitive Walkthrough**  
**Review-based Evaluation.**





# Expert Analysis

- Expert analysis: designer or HCI expert assesses a design based on known/standard cognitive principles or empirical results.
  - Expert analysis methods can be used at any stage in the life cycle.
  - Expert analysis methods are relatively cheap.
  - Expert analysis methods, however, do not assess the actual use of the system.
- Examples of expert analysis methods: Heuristic Evaluation (HE), Cognitive Walkthrough (CW), Review-based Evaluation.



---

# Heuristic Based Evaluation Technique

**(“to discover”)** pertains to the process of gaining knowledge or some desired result by intelligent guesswork rather than by following some pre established formula.

---





# Introduction

- Heuristic Evaluation (HE) was proposed by Nielsen and Molich. ( **Read Nielsen's ten Heuristics already discussed earlier**)
  - In HE, experts scrutinize the interface and its elements against established usability heuristics [*another previous tutorial*].
  - The experts should have some background knowledge or experience in HCI design and usability evaluation.
-





## The Process..

- 3 to 5 experts are considered to be sufficient to detect most of the usability problems.
- The enlisted experts are provided with the proper roles (and sometimes scenarios to use) to support them when interacting with the system/prototype under evaluation.
- They then evaluate the system/prototype individually. This is to ensure an independent and unbiased evaluation by each expert.
- They assess the user interface as a whole and also the individual user interface elements. The assessment is performed with reference to a set of established usability principles.
- When all the experts are through with the assessment, they come together and compare and appropriately aggregate their findings.



# Cognitive Walkthrough

**Relating to the mental processes of perception, memory, judgment, and reasoning, as contrasted with emotional and volitional processes.**

---



# Introduction

- Cognitive Walkthrough (CW) was proposed by Polson et al.
  - CW evaluates design on how well the design supports user in learning the task to be performed [primarily through exploration i.e. hands on].
  - •CW is usually performed by expert in cognitive psychology.
-



## The Process..

- The expert 'walks through' the design [i.e. steps through each step of some known/representative task] to identify potential problems.
  - 4 requirements in order to perform the CW:
    1. specification or prototype of the system
    2. description of the task the user is to perform
    3. complete, written list of actions constituting the task
    4. description of the user (including the level of experience and knowledge)
-



## The Process..

- With the foregoing information, the evaluator steps through each of the actions trying to answer the following 4 questions:
  1. Is the *effect of the action the same as the user's goal at that point? [what the action will do/action's effect should be what the user intends/user's goal.]*
  2. will users see that the action is *available [when they want it] - visibility at that time?*
  3. once users have found the correct action [as in the foregoing], will they *know/recognize it is the one they need? [effective representation of the action, clear representation.]*
  4. after the action is taken, will users understand the *feedback they get? [effective confirmation that the action has been taken.]*





## The Process..

- forms are used to guide analysis e.g.
  - cover form [for the four requirements above, date, time, evaluators of the CW],
  - answer form [for answering the four questions above],
  - usability problem report [for describing any negative answers/problems, severity of the problem e.g. frequency of occurrence and seriousness of the problem, date, time, evaluators].
-



# Review Based Evaluation

---





# Introduction

- Experimental results and empirical evidence from the literature [e.g., from psychology, HCI, etc] can be used to support or refute parts of design.



## The Process..

- It is expensive to repeat experiments continually and therefore a review of relevant literature can save resources (e.g., effort, time, finances, etc).
  - However, care should be taken to ensure results are transferable to the new design
  - [e.g., note the design in consideration, the user audience, the assumptions made, etc].
-



# Model Based Evaluation



# How to Proceed!

- Cognitive models can be used to filter design options
  - e.g. GOMS (Goals, Operators, Methods and Selection) model can be used to predict user performance with a user interface, keystroke-level model can be used to predict performance for low-level tasks
-



# User Based Evaluation



Hello Participants!!

---



- User-based evaluation basically is evaluation through user participation i.e. evaluation that involves the people for whom the system is intended; the users.
-



## The Process

- User-based evaluation techniques include:
    - experimental methods,
    - observational methods,
    - query techniques (e.g., questionnaires and interviews),
    - physiological monitoring methods (e.g., eye tracking, measuring skin conductance, measuring heart rate).
  - User-based methods can be conducted in the laboratory and/or in the field
-





## Using Laboratory

### □ **Advantages:**

- Specialist equipment available.
- Uninterrupted environment.

### □ **Disadvantages:**

- Lack of context.
- Difficult to observe several users cooperating.

### □ **Appropriate:**

- If system usage location is dangerous, remote or impractical.



## Field or Working Environment

### □ **Advantages:**

- Natural environment.
- Context retained (though observation may alter it).
- Longitudinal studies possible.

### □ **Disadvantages:**

- Field challenges e.g., distractions, interruptions, movements, danger, noise.

### □ **Appropriate:**

- Where context is crucial
-



# Query Technique

**Asking questions**

---



# Observational Methods

---



# Introduction

- Observational methods
    - think aloud,
    - cooperative evaluation,
    - protocol analysis,
    - post-task walkthroughs
    - Psychological Monitoring
-



# Think aloud

- User is observed performing task.
  - User is asked to describe what s/he is doing and why, what s/he thinks is happening, etc.
  - **Advantages:**
    - Simplicity - requires little expertise.
    - Can provide useful insight.
    - Can show how system is actually used.
  - **Disadvantages:**
    - Subjective [really depends on the user].
    - Selective [out of many things, the user may choose what to describe].
- 
- Act of describing may alter task performance.



---

# Cooperative Evaluation

- Variation on think aloud.
  - User collaborates in evaluation.
  - Both user and evaluator can ask each other questions throughout.
  - **Additional advantages:**
    - Less constrained and easier to use.
    - User is encouraged to criticize system.
    - Clarification possible.
-





# Protocol Analysis

- Paper and pencil: cheap, limited to writing speed.
  - Audio: good for think aloud, difficult to record sufficient information to identify exact actions in later analysis, difficult to match with other protocols ('synchronization').
  - Video: accurate and realistic, needs special equipment, obtrusive.
  - Computer logging: automatic and unobtrusive, large amounts of data difficult to analyze.
  - User notebooks: coarse and subjective, useful insights, good for longitudinal studies.
- Mixed use in practice.
- Audio/video transcription difficult and requires skill.
- Some automatic support tools available e.g., EVA (Experimental Video Annotator), Observer Pro (from Noldus), Workplace project (Xerox PARC), etc.



## Post-task Walkthrough

- Transcript played back to participant for comment i.e. user reacts on action after the event.
- Used to fill in intention i.e. reasons for actions performed and alternatives considered.
- It also is necessary where think aloud is not possible.
- **Advantages:**
  - Analyst has time to focus on relevant incidents.
  - Avoids excessive interruption of task.
- **Disadvantages:**
  - Lack of freshness.
  - May be post-hoc interpretation of events.



# Physiological monitoring methods

- [e.g., eye tracking, measuring skin conductance, measuring heart rate].
- **Eye-tracking**
  - Head or desk mounted equipment tracks the position of the eye.
  - Eye movement reflects the amount of cognitive processing a display requires.
  - Measurements include: fixations, scan paths, etc. For instance:
    - number of fixations.
    - duration of fixation.
- scan paths: moving straight to a target with a short fixation at the target is optimal.



# Psychological Measurements

## □ *Emotional response linked to physical changes.*

- These may help determine a user's reaction to a user interface.
- Measurements include: heart, sweat, muscle, brain. For instance:
  - *heart activity: e.g. blood pressure, volume and pulse.*
  - *activity of sweat glands: Galvanic Skin Response (GSR).*
  - *electrical activity in muscle: electromyogram (EMG).*
  - *electrical activity in brain: electroencephalogram (EEG).*
  - There is some difficulty in interpreting these physiological responses; more research is needed.

## Types of Evaluation Methods : Cheat Sheet

	Think Aloud Protocol	Heuristic Evaluation	Cognitive Walkthrough	Model Human Processor	GOMS	Experiment	Competitive Analysis
User	X					X	
Heuristic		X	X				
Hard-Data				X	X	X	
Market Defined							X
Speed	Fast	Medium	Medium/Slow	Slow	Slow	Slow	Fast
Cost	Cheap	Cheap	Cheap	Cheap	Cheap	Expensive	Cheap
# of users required	2-8	0	0	0	0	20+	0
# of evaluators required	1-2	2-8	2-8	1-2	1-2	1-4	1-2
Developed by	Newell & Simon from CMU	Jakob Nielsen (useit.com)	Based on Lewis & Polson's CE+ info processing model	Card, Moran & Newell	Card, Moran & Newell	Social Psychology as a field	Business & Marketing as an org structure





---

## Factors that can influence the choice

- when in process: design vs. implementation
  - style of evaluation: laboratory vs. field
  - how objective: subjective vs. objective
  - type of measures: qualitative vs. quantitative
  - level of information: high level vs. low level
  - level of interference: obtrusive vs. unobtrusive
  - resources available: time, subjects, equipment, expertise
-