

**HUMAN-COMPUTER INTERACTION** **THIRD EDITION** **DIX FINLAY ABOUW BEALE**

chapter 9

evaluation techniques

**HUMAN-COMPUTER INTERACTION**

### Evaluation Techniques

- Evaluation
  - tests usability and functionality of system
  - occurs in laboratory, field and/or in collaboration with users
  - evaluates both design and implementation
  - should be considered at all stages in the design life cycle

**HUMAN-COMPUTER INTERACTION**

### Goals of Evaluation

- assess extent of system functionality
- assess effect of interface on user
- identify specific problems

**HUMAN-COMPUTER INTERACTION**

### Evaluating Designs

Cognitive Walkthrough  
Heuristic Evaluation  
Review-based evaluation

**HUMAN-COMPUTER INTERACTION**

### Cognitive Walkthrough

Proposed by Polson *et al.*

- evaluates design on how well it supports user in learning task
- usually performed by expert in cognitive psychology
- expert 'walks through' design to identify potential problems using psychological principles
- forms used to guide analysis

**HUMAN-COMPUTER INTERACTION**

### Cognitive Walkthrough (cont)

- For each task walkthrough considers
  - what impact will interaction have on user?
  - what cognitive processes are required?
  - what learning problems may occur?
- Analysis focuses on goals and knowledge: does the design lead the user to generate the correct goals?

## Cognitive Walkthrough (cont)

Walkthrough requires:

- A detailed specification or prototype of the system
- A description of a representative task
- A list of the actions needed to complete the task
- Users' experience and knowledge

## Cognitive Walkthrough (cont)

For each step in the action sequence, the evaluator try to answer 4 questions:

- Is the effect of the action the same as the user's goal at that point?
- Will user see the action as available?
- Once users found the correct action, will they know it is the one they need?
- After the action is taken, will users understand the feedback they get?

## Cognitive Walkthrough (cont)

Cognitive Walkthrough For each action:

Cover Sheet

- Interface:
- User(s):
- Evaluators:
- Date:
- Task:
- Action sequence:
- User's initial goals:

1. Action->goal?
2. Action is available?
3. Is the action the one which is needed?
4. Is the feedback clear?

## Heuristic Evaluation

- Proposed by Nielsen and Molich.
- usability criteria (heuristics) are identified
- design examined by experts to see if these are violated
- Example heuristics
  - system behaviour is predictable
  - system behaviour is consistent
  - feedback is provided
- Heuristic evaluation 'debugs' design.

## Heuristic Evaluation

The severity of each problem is also assessed along 4 factors:

- How common is the problem
- How easy is it for user to overcome
- Is it a one-off or a persistent problem
- How seriously will the problem be perceived

Severity rating on a scale

- 0 = no usability problem
- 1 = cosmetic problem, address if time
- 2 = minor problem, low priority
- 3 = major problem, high priority
- 4 = catastrophe, imperative to fix.

## Nielsen's 10 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 recognise, diagnose and recover from errors
10. Help and documentation

## Review-based evaluation

- Results from the literature used to support or refute parts of design.
- Care needed to ensure results are transferable to new design.
- Model-based evaluation
- Cognitive models used to filter design options  
e.g. GOMS prediction of user performance.
- Design rationale can also provide useful evaluation information

## Evaluating through user Participation

## Laboratory studies

- Advantages:
  - specialist equipment available
  - uninterrupted environment
- Disadvantages:
  - lack of context
  - difficult to observe several users cooperating
- Appropriate
  - if system location is dangerous or impractical for constrained single user systems to allow controlled manipulation of use

## Field Studies

- Advantages:
  - natural environment
  - context retained (though observation may alter it)
  - longitudinal studies possible
- Disadvantages:
  - distractions
  - noise
- Appropriate
  - where context is crucial for longitudinal studies

## Evaluating Implementations

Requires an artefact:  
simulation, prototype,  
full implementation

## Experimental evaluation

- controlled evaluation of specific aspects of interactive behaviour
- evaluator chooses hypothesis to be tested
- a number of experimental conditions are considered which differ only in the value of some controlled variable.
- changes in behavioural measure are attributed to different conditions

HUMAN-COMPUTER INTERACTION

## Experimental factors

- Subjects
  - who – representative, sufficient sample
- Variables
  - things to modify and measure
- Hypothesis
  - what you'd like to show
- Experimental design
  - how you are going to do it

HUMAN-COMPUTER INTERACTION

## Variables

- **Independent variable (IV)**
  - the one we think **causes** a change
  - is directly changed or **manipulated**
  - characteristic changed to produce different conditions  
e.g. interface style, number of menu items
  - condition
    - a specific amount of independent variable
    - level or treatment
- **Dependent variable (DV)**
  - characteristics measured in the experiment  
e.g. time taken, number of errors.
  - the one we think is the **effect**, the outcome
  - is **only measured** or registered - not manipulated

HUMAN-COMPUTER INTERACTION

## Hypotheses

Scientific hypothesis consists of two separate mutually exclusive hypotheses:

**Alternative Hypothesis**  
A specific statement of prediction that usually states *what we expect will happen* in our study

**Null Hypothesis**  
The hypothesis that describes the *possible outcomes other* than the alternative hypothesis. Usually it predicts there will be *no effect* of a program or treatment we are studying.

HUMAN-COMPUTER INTERACTION

## Hypothesis

- prediction of outcome
  - framed in terms of IV and DV
  - e.g. "error rate will increase as font size decreases"
- null hypothesis:
  - states no difference between conditions
  - aim is to disprove this
  - e.g. null hyp. = "no change with font size"

HUMAN-COMPUTER INTERACTION

## Hypothesis

- An idea on trial.
- Relationship between at least two variables.
- A specific statement of prediction

Question → Hypothesis  
Answer ← Testing

```

graph TD
    Hypoth --> TEST
    TEST -- Yes --> extend["(extend)"]
    TEST -- No --> revise["(revise)"]
    extend --> NewHypoth
    revise --> NewHypoth
  
```

HUMAN-COMPUTER INTERACTION

## Hypothesis

<p><b>Criteria</b></p> <ul style="list-style-type: none"> <li>Testable</li> <li>Falsifiable</li> <li>Precise</li> <li>Rational</li> <li>Parsimonious</li> </ul>	<p><b>Questions about hypothesis</b></p> <ul style="list-style-type: none"> <li>Can a test be designed for it?</li> <li>Can it be possibly proved false?</li> <li>Are its items clearly defined?</li> <li>Does it fit with the known information?</li> <li>Does it involve the simplest possible approach?</li> </ul>
---	---

**Experimental Design**

- within groups design
  - each subject performs experiment under each condition.
  - transfer of learning possible
  - less costly and less likely to suffer from user variation.
- between groups design
  - each subject performs under only one condition
  - no transfer of learning
  - more users required
  - variation can bias results.

**Experiment Design**

- Testing causal hypothesis
- What do we need for proof?
  - Manipulation
  - Control
  - Random assignment

**Experimental Design**

**Control**

- Control group
  - a group that receives the normal amount of the independent variable (usually zero).
- Experimental group
  - a group that receives a certain amount of the independent variable (usually nonzero) and so do experiences the treatment









**Random assignment**

- groups comparability

**Example - Experiment**

- Research purpose: to study the icon style for an interface of a document-processing package and its impact on recall.
- Hypothesis: naturalistic icons will be easier to recall
- Dependent variable: **accurate recall, speed of recall**
- Independent variable: **icon style: naturalistic and abstract**

**Experimental Design**

	IV	DV	IV	DV
G1				
	Naturalistic icon		Abstract icon	
G2				
	Abstract icon		Naturalistic icon	