

CMC092/ITC082/CMC262 Human-Computer Interaction

Lecture 2 Evaluation

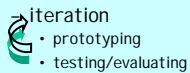
Evaluation

- **why** evaluation?
- **what** do we evaluate?
- **who** evaluates?
- **how** do we evaluate?

..

Why evaluation?

- concerned here with *usability* not functionality
- (but can they be distinguished?)
- evaluation is **crucial**
- successful (good, usable) design requires



..

Why test for usability?

- usability is **not**
 - predictable
 - intuitive, 'common sense'
- designers are **not**
 - typical
 - typical people? typical users?
- the technical tendency
 - *feature creep*
 - engineers see more functionality
 - users see more complexity

..

So... evaluation & testing

- is required
- must be
 - valid, reliable, informative
 - based on evidence
- documented with suitable
 - usability *metrics*

..

What do we evaluate?

- *usable* means...?
 - easy to use?
 - easy to learn?
 - intuitive?
 - enjoyable?
 - transparent?
 - fast?
 - reliable?

..

All aspects of the interface

- i/o devices
- screen layout and look
 - colour, clutter, position
- access to functionality
 - commands, menus
- navigation and sign-posting
- error-handling
- documentation and help systems

..

Who evaluates?

- academic researchers
- system designers
- hci 'experts'
- **users, users, users**

..

How do we evaluate?

- *evaluation* is a broad term
- a range of activities, objectives inc:
- academic/research activities
 - controlled experiment
- evaluation for system design and development
 - usability testing
- user surveys etc

..

Evaluation & testing

- for each kind of evaluation, at each stage, appropriate tests can be devised
- 2 goals of testing (Johnson)
 - informational
 - social

..

Testing may be

- formative evaluation
 - before implementation
 - part of design process
 - prototyping & testing
 - informs improvements
- summative evaluation
 - comparative

..

Evaluation at all stages in the system development cycle

- at analysis
- in design
- in development
- on completion

..

Evaluation methods

- analytical
 - expert analysis
- empirical
 - observational/ethnographic
 - survey
- experimental
- co-operative evaluation

..

Analytical evaluation

- carried out by the HCI specialist/system developer
- formal analysis of task
 - various methodologies
 - GOMS
 - task analysis

..

Expert evaluation

- 'expert' performs cognitive walkthrough
- 'expert' here means HCI expert not domain expert
- identify problems based on
 - human factors experience
 - interface experience

..

Applications of analytic evaluation

- good for comparison of different systems
 - alternative implementations
 - different software choices
- predictions of user performance useful in
 - usability specifications
 - development of usability metrics
- limitations
 - cannot substitute for real testing by real users

..

Empirical evaluation

- real data about real people using the interface
- in the normal working environment or
- in formal test environment
- techniques
 - observation
 - survey

..

Ethics in testing and experimentation

- issues for test subjects
- informed consent
 - purpose and means of evaluation
 - preparation for test responses/stress
- awareness of emotional response
- freedom to stop
- confidentiality
- issues apply also to software instrumentation

..

Recap – expert evaluation (without 'real people')

- cognitive walkthrough
 - walk through tasks/functions/use cases
 - use experience, heuristics to identify problem areas
- analyse using formal analytic methodologies eg
 - GOMS
 - hierarchical task analysis

..

Expert evaluation good for

- informing design decisions
- informing prototype decisions
- avoiding well-recognised problems
- application of basic guide-lines
- informing test requirements
- does not substitute for evaluation by users

..

Evaluation involving users

- where possible, should be real end-users of the system being evaluated
- may take place within normal work - place/time
 - tasks involved would be those of normal working practice
- outside normal work
 - selected subjects, selected/devised tasks

..

Choosing evaluation tasks and subjects

- often determined by the requirements of the evaluation
- in general, tasks should be representative of the real-world system applications
- users should be representative of the real system users
 - but - Exactly who are the users?

..

Methods include

- observing users
 - in the work-place, in normal use context
 - ethnographic studies
 - in controlled/special environment
- survey user opinion/experience
 - interview
 - questionnaire

..

Ethnographic studies

- ethnomethodologies
 - makes no assumptions about human behaviour or methods
 - purely observational/experiential
- researcher enters the real-world domain
 - observation of
 - system tasks & interactions
 - collaborative interactions
 - workflow

..

Observation techniques

- physical observation
 - researcher may or may not be present
 - subject system interaction is recorded
 - film, audio, notes
 - may also be system instrumentation
 - use of 'think-aloud'
 - observation labs
 - hci research centres
 - users observed less obtrusively

..

Observation techniques (2)

- collaborative evaluation
 - user in dialogue with developer
 - walk-through tasks
 - discuss issues/experience
- normal in participative design approaches

..

Problems with observation

- observation influences user behaviour
 - physical presence
 - mechanical presence
 - lab environment
- not normal working environment
- task selection?

..

Further issues with observation

- volume of data generated
- the more material the more to
 - transcribe
 - evaluate
 - analyse

..

Software instrumentation

- system records user interactions
- can report on
 - error rates
 - keystrokes
 - time to complete task
 - use of on-line help
 - repeated patterns of user actions
 - indicate optimisation possibilities

..

Survey evaluation

- interviews
 - time-consuming
 - allow for in-depth exploration of opinions/attitudes
- questionnaires
 - allow for large numbers to be surveyed
- both need careful design (more next week)

..

Experimental evaluation

- HCI academic research activity
- construction of experiment
 - test specific hypothesis
 - provide quantitative data
- requires resources to construct formal experiments in test conditions

..

What do we evaluate?

- respond to problems?
- form hypothesis?
- formulate process to test hypothesis?
- devise test
 - standard tasks?
 - design survey?
 - metrics?
- subjects and controls?

..

How do we measure what we are evaluating? (Metrics)

- time to complete task?
- time to learn a process?
- number of keystrokes/mouse clicks required?
- ratings on an attitude scale?
- We need some way to quantify, objectively measure

..

Selection/design of tasks

- representative, real-world tasks
- use cases, scenarios
- afford evaluation of the things you want to discover?

..

Selection/design of tasks (2)

- What question are we trying to answer?
- What data will let us answer that question?
- How can we get the data?

..