DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF TORONTO

CSC 318S

THE DESIGN OF
INTERACTIVE COMPUTATIONAL MEDIA

Lecture 14 — 9 March 1998

SYSTEM AND INTERFACE EVALUATION 1

Ronald Baecker
Professor of Computer Science,
Electrical and Computer Engineering, and Management
University of Toronto

_____

14.1 A framework of empirical evaluation

Goals of empirical evaluation

    Understanding what happens in practice
        Interface quality
        Interface problems
        What factors seem to affect interface quality —
            defined very broadly, to include metaphors,
            mental models, look & feel, documentation, etc.)

    Understanding why it happens
        Often users will tell us why, or we can deduce it
        from observations and analysis
        Yet there are limits to empirical approach
        We also build models to achieve a deeper
        understanding of casual factors (CSC428)
        Since there is little compelling evidence that this is
            generally cost-effective, we usually rely on
        empirical approaches

Research approaches — four dimensions (Ray & Ravizza, 1985)
    Methodology
        Naturalistic observation (watching & recording)
        True experiments (manipulating & measuring)
    Setting — Field research or laboratory research
    Experimental role — Scientist as participant or observer
    Size: one, few, or many subjects

Naturalistic observation
    Noninterference with phenomena
    Observations of patterns and invariants
    Useful for big picture
    Qualitative, descriptive (typically)

_____

True experiments
     Interfere, manipulate, control
     Measurements of observed patterns
     Detailed results
     Quantitative

A deeper approach is that of McGrath... (see Ch. 2 in BGBG)...
     but first lets look at the relationship of evaluation to design

14.2 Empirical evaluation in design

Fig. 14.1. Typical design/development process (BGBG, Fig. 2.1 revised, p. 74)

	Design-->	Implement-->	Evaluate-->
Information Collection	Questionnaires Interviews with discipline specialists Characterizations of users and tasks	"Day in the life" scenarios	e.g., interviews with users to get reactions to "Day in the life" scenarios
Concept design	Initial design concepts	Design mockups	e.g., interviews with users to get reactions to design mockups
Functionality and interface design	Design of system functionality and look-and-feel	Implementation of "smoke and mirrors" prototype	e.g., usability tests
Prototype implemen-tation	Design of "critical mass" of system	Implementation of partial working system	e.g., usability inspection
Deliverable system im-plementation	Design and modification of deliverable system	Implementation and installation of this system	e.g., usability tests
System en-hancement and evolution	Design of monitoring and feedback system	Implementation of this system	e.g., interviews & questionnaires

_____

14.3 HCI research and evaluation strategies

Fig. 14.2 lists some strategies, and which we shall discuss. In Lect. 15 we organize them according to a taxonomy from McGrath, and position them in the development cycle

Fig. 14.2. HCI research & evaluation methods (BGBG, Fig. 2.5 revised, p. 81)

Field strategies
  (Observations, with settings under conditions as natural as possible)
Field studies
        Ethnography & interaction analysis (This lecture, more in CSC428)
        Contextual inquiry (CSC428)
Field experiments
        Beta testing of products (CSC 454)
        Studies of technological change (CSC 300)

Experimental strategies
  (Experiments, in settings concocted for research purposes)
Experimental simulations
        User testing or Usability testing (This lecture, more in CSC428)
        Usability engineering (CSC428)
Laboratory experiments
        Controlled experiments (A little in Lecture 15, more in CSC428)

Respondent strategies
  (Asking questions, where setting is muted or made moot)
Judgment studies
        Demonstrations (Lecture 15)
        Usability inspection methods (A little in Lecture 15, more in CSC428)
        Cognitive walkthroughs (CSC428)
Sample surveys
        Questionnaires, surveys, & interviews (This lecture, more in CSC428)

Theoretical strategies
  (No observation of behavior, experiments, or questions required)
Formal theory
        Design theory, e.g., Norman's 7 Stages (CSC428)
        Behavioral theory, e.g., color vision (CSC428)
Computer simulations
        Human information processing theory (CSC428)

_____

14.4 Questionnaires, surveys, and interviews

Methods
- *Questionnaire* design
  - Formulating & asking questions, & analyzing answers
  - Hard to avoid bias in the phrasing of questions
  - Therefore requires pre-testing ("pilot testing")
- *Surveys* — (possibly large-scale) administration of questionnaires to appropriate *samples* of individuals chosen from a *population*
- Administration of questions through *interviews*

Roles in information gathering
- Understanding user needs
- Carrying out a task analysis
- Understanding the work context

Advantages
- Easy to administer, efficient time-wise
- Compared to surveys, interviews can be free-ranging, adapting to conversation flow, yielding novel insights

Disadvantages
- Subjects may not be able to conceptualize new technology
- Abstract, may not be well grounded in real application
- Danger of bias – Putting words into people's mouths

Examples
- Gould & Boies interviews with principals that led to design of Speech Filing System (B&B, pp. 8-37)
- Posner's interviews w. collaborative writers (Baecker et al., BGBG, 775-782)

See also notes from Lecture 6

_____

14.5 Scenarios, mockups, prototyping, manuals

Method
>Produce facsimile of interface
>>Construct scenario — words, pictures, animation, video
>>Build physical mockups
>>"Program" interactive prototypes
>>Write manual in advance of system
>Design and produce situation
>Observe behaviour of users

Roles
>Elicits initial reactions to, problems with user's model,
>>interface
>Engages potential users, gets them excited and
>>involved, makes them see that they can contribute

Advantages
>Can be done in advance of actual system construction

Disadvantages
>Facsimile will not embody all characteristics of system

Examples
>Scenarios, prototypes in CHI'89 Information Kiosk
>>(Salomon, BGBG, Case A, 23-34)
>Wizard of Oz simulation of a listening typewriter
>>(Newell, et al.)

_____

14.6 User testing

Methods
>   Design scenario
>   Record user behaviour
>> Typical usage, or critical incidents
>> Keystroke recording (14.8)
>> Thinking aloud protocols (14.7)
>> Videotape protocols (14.7, 14.8)
>   Interviews for subjective impressions
>   Analyze user behaviour
>> Protocol analysis (14.8)

Roles
>   Understanding user methods
>   Understanding user problems
>   Discovering user thought processes

Advantages
>   Observation of real usage of real system

Disadvantages
>   Individuals often know only limited aspects of a system

Examples
>   Mack, et al. studies of text editor learning (BGBG, Ch. 10)
>   Usability tests of SASE, SASSE (BGBG, Ch. 11)

_____

Typical steps in user testing – Gomoll (Laurel, pp. 85-90)
    1. Set up the observation
    2. Describe the purpose of the observation
    3. Tell the user that it's OK to quit at any time
    4. Talk about & demonstrate the equipment in the room
    5. Explain how to "think aloud"
    6. Explain that you will not provide help
    7. Describe the task and introduce the system
    8. Ask if there are any questions before you start;
       then begin the observation
    9. Conclude the observation & debrief (interview) subjects
    10. Analyze the data
    11. Tabulate the results
    12. Interpret the results in the context of other results

See Figure 14.3 for more detailed procedure

14.7 Thinking aloud

Attempt to elicit thought processes of subject

Subject talking while they are doing
    Problems they are having
    Solutions they are considering
    Why they are having trouble

Pairs of subjects conversing (Co-Discovery Learning,
    Kennedy paper in BGBG, pp. 182-185)

Slows down the process, but yields valuable insights

_____

Figure 14.3 User testing with thinking aloud
(BGBG, Fig. 2.8, p. 85, adapted from Neilsen, 1992a)

Practical study design
• Reflect on the participants' backgrounds and how they might affect the study
• Be aware of problems that arise when experimenters know the users personally
• Prepare for the study carefully (avoid last minute panic)
• Select the tasks carefully to be representative and to fit the allotted time
• In general, start with an easier (but not frivolous) task
• Write down features of the system that are *not* being tested as well as those that are!
• Define the start-up state for the study precisely
• Define precise rules for when and how users can be helped during the study
• Plan the timing and cut-off procedure (if subject gets stuck) for each part of the study
• Include reasonable provisions for data collection (e.g., audio, video, or keystroke capture
   where appropriate)
• Plan data analysis techniques in advance
• Carry out a pilot study (important but often overlooked)

Written materials
• Participant release form
• Questionnaire covering prior experience etc. (if relevant)
• Introduction to the study for users, including scenario of use
• Checklist for experimenters
• Evaluation survey (if relevant)

Carrying out the study
• Let users know that complete anonymity will be preserved
• Let them know that they may quit at any time
• Stress that the system is being tested, not the participant
• Indicate that you are only interested in their thoughts relevant to the system
• Demonstrate the thinking-aloud method by acting it out for a simple task, e.g., figuring out
   how to load a stapler
• Hand out instructions for each part of the study individually, not all at once
 •Maintain a relaxed environment free of interruptions
• Occasionally encourage users to talk if they grow silent
• If users ask questions, try to get them to talk (e.g., "What do you think is going on?" and
   follow predefined rules on when to help or interrupt to help.
• Debrief each user after the experiment

Improving the study
• The pilot study should "debug" the study. This minimize changes during the study, allowing
   quantitative data analysis. But improvements may be warranted.
• Experimenters' role can be improved
• Tasks given to participant can be improved
• Written materials can be improved

_____

14.8 Data capture, video taping, and protocol analysis

Keystroke logging
    Sometimes known as "dribble files"
    Record precise user behaviour
    Record times to carry out actions
    Record user errors

Observation and notetaking
    Critical incidents
    User problems

Audio and video recordings
    Can't record all behaviour in real-time
    Preserve behaviour for review
    Non-verbal behaviour
    Behaviour in context

Data extraction and analysis
    Enormous amounts of "data" in the video tape
    How to convert to information?
    Review of video and audio tapes
    Construction of *protocols*
    Classifying events and counting events
    Tools for video annotation and analysis

Ethnographic methods, interaction analysis (Suchman & Trygg)
    Ethnography..."the careful study of activities and
        relations between them in a complex *social setting*."
    Situations of use in actual work settings
    Interaction analysis..."to uncover the regularity and
        efficacy of peoples' relations with each other and their
        use of the resources that their environment affords"
    See also notes from Lecture 6