HCI Application: Qualitative Evaluation

Organizational & Social Issues

Task

Design

Technology
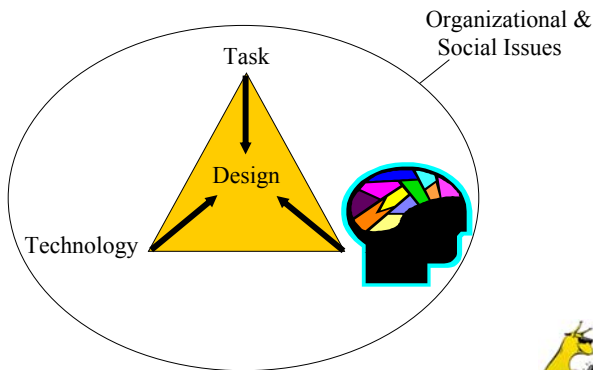
Why Do We Evaluate?

A. Pre-design
- what do people do?
- how can we understand what we need in system functionality?

► Evaluation produces
- key tasks, functional and non functional requirements
- work practices
- organizational practices
- user types

Why Do We Evaluate?

B. During initial stage, developing design ideas and representations
- evaluate choices of initial design ideas and representations
  - ► is the representation appropriate?
  - ► does it reflect how people think of their task?

► Evaluation produces:
- user reaction to high-level design
- validation and list of problem areas at conceptual level
- new design ideas

Why Do We Evaluate?

C. During iterative development, refining a design / representation
- fine tune the interface, looking for usability bugs
  - ► can people use this system?

► Evaluation produces:
- user reaction to low-level design
- validation and list of problem areas (bugs)
- variations in design ideas

Why Do We Evaluate?

D. Post-design
- *acceptance test:* did we deliver what we said we would?
- *revisions:* what do we need to change?
- *effects:* What did we change in the way people do their tasks?

► Evaluation produces
- testable usability metrics
- actual reactions
- validation and list of problem areas (bugs)
- changes in original work practices/requirements

Evaluation Process

1. Developing a test plan ("blueprint" for the test)
- General intention and individual checks
  - ► MUST BE PRECISE
  - ► MUST BE CLEAR
  - ► MUST BE MEASURABLE/OBSERVABLE
- Specifies the resources needed to carry out test
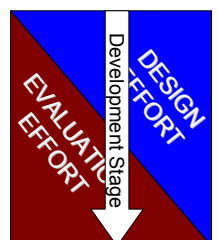- Governs conclusions that can be made
2. Developing test materials
3. Recruiting participants
4. Carrying out the test
5. Debriefing participants
6. Findings and recommendations

DESIGN EFFORT
Development Stage
EVALUATION EFFORT

Test plan

1. Purpose of test
2. Problem statements/test objectives
3. Participant profile (inclusion/exclusion criteria)
4. Method/technique to be used
5. List of tasks to be used
6. Test environment (field vs. lab) and material (HW/SW, resources → recorder and batteries, report forms, questionnaires)
7. Experimenter's role (monitor, coach etc.)
8. Evaluation measures to be taken (qualitative vs. quantitative, subjective vs. objective)
9. Contents of report to be produced and how the report is going to be presented → focus group, informal meeting, big bosses are going to be there

Quantitative vs. Qualitative

Quantitative	Qualitative
Numbers	Words
Point of view of researcher	Points of view of participants
Researcher distant	Researcher close
Theory testing	Theory emergent
Static	Process
Structured	Unstructured
Generalization	Contextual understanding
Hard, verifiable data	Rich, deep data
Macro	Micro
Behavior	Meaning
Artificial settings	Natural settings

Evaluation measures

	Quantitative	Qualitative
Subjective	How many errors do you think you made during the test?	Do you believe you did well or poorly, overall? Did you feel stressed during the test?
Objective	How many errors were made during the test?	Was the participant's performance: • Poor • Adequate • Good • Excellent Did the participant shake or sweat during the test?

Why Use Different Methods?

► Information requirements differ
  ▪ pre-design, iterative design, post-design, generalizable knowledge…
► Information produced differs
  ▪ outputs should match the particular problem/needs
► Cost/benefit of using a certain method
► One method's strength can complement another's weakness
  ▪ no one method can address all situations
► Constraints
  ▪ may force you to chose quick and dirty discount usability methods

How Can We Compare Methods?

► Type of information (qualitative vs. quantitative)
► Relevance
  ▪ does the method provide information to our question / problem?
► Setting
  ▪ is it important that the system be evaluated in-context?
► Generalization
  ▪ how well can I generalize the information produced to other situations?
► Repeatability
  ▪ would the same results be achieved if the test were repeated?

How Can We Compare Methods?

► Quickness
  ▪ can I do a good job with this method within my time constraints?
► Cost
  ▪ is the cost of using this method reasonable for my question?
► Equipment
  ▪ What special equipment / resources required?
► Personnel, training and expertise
  ▪ What people / expertise are required to run this method?
► Validity
  ▪ External validity: can the results be applied to other situations?
  ▪ Internal validity: do we have confidence in our explanation?

How Can We Compare Methods?

► Subject selection
  ▪ how many do I need, who are they, and can I get them?
► Scope of subjects
  ▪ is it good for analyzing individuals? small groups? organizations?
► Control
  ▪ do I need to control for certain factors to see what effects they have?
► Cross-sectional or longitudinal
  ▪ is it important that changes over time are measured?
► Support
  ▪ are there tools for supporting the method and analyzing the data?
► Comparative
  ▪ can I use it to compare different things?

How Can We Compare Methods?

► Does the test measure something of relevance to usability of real products in real use outside of lab?
  ▪ Some typical reliability problems of testing vs. real use
    ► non-typical users tested
    ► tasks are not typical tasks
    ► physical environment different
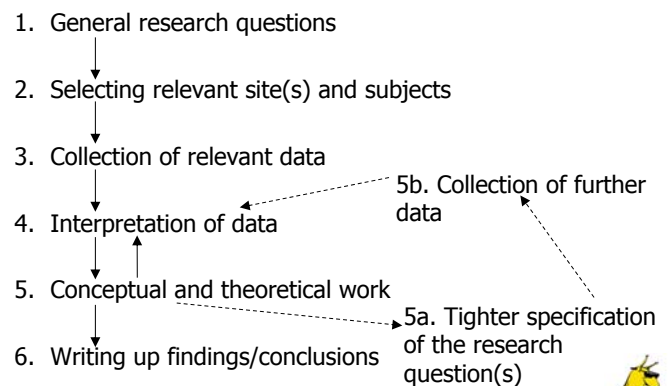   ► social influences different

---

Qualitative Research

► Main research methods:
  ▪ Ethnography/participant observation
  ▪ Qualitative interviewing
  ▪ Focus groups
  ▪ Language-based approaches: conversation analysis; discourse analysis
  ▪ Collection and qualitative analysis of texts and documents

---

Main Steps in Qualitative Research

1. General research questions
2. Selecting relevant site(s) and subjects
3. Collection of relevant data
4. Interpretation of data
5. Conceptual and theoretical work
6. Writing up findings/conclusions

5b. Collection of further data

5a. Tighter specification of the research question(s)

The Qualitative Interview

► Structured interview
  ▪ Replicable but may lack richness
  ▪ Questions with a set of pre-defined answers
  ▪ Order and wording are important
  ▪ Surveys
► Open Qualitative interviews
  ▪ Qualitative analysis
  ▪ Questions are used to start conversations with informants
  ▪ Unstructured - are not directed by a script. Rich but not replicable
  ▪ Semi-structured - guided by a script but interesting issues can be explored in more depth. Can provide a good balance between richness and replicability

---

The Qualitative Interview

Qualitative interviewing	Structured interview
much less structured	structured for reliability and validity of measurement of key concepts
greater interest in the interviewee's point of view	interview reflects the researcher's concerns
`rambling' is often encouraged	rambling regarded as a nuisance and discouraged
interviewers can depart significantly from any schedule or guide	any departure compromises the standardization of the interview process
flexible, responding to the direction interviewees take the interview	typically inflexible, because of the need to standardize
researcher wants rich, detailed answers	interview supposed to generate answers that can be coded and processed quickly
the interviewee may be interviewed 1+	unless longitudinal study, the person will be interviewed 1x

Interview Guide

► A brief list of memory prompts of areas to be covered
► A more structured list of issues to be addressed or questions to be asked
► A list of themes and probing questions
► Preparing interview guide:
- Prepare enough number of question
- Put checkbox in case you need to deviate to follow user lead
- Check whether the topics cover all of your research questions
- Use language that is comprehensible and relevant to the people you are interviewing
- Record `fact sheet' information →general (age, gender, etc.) and specific (position in company, Internet exp, etc.) to contextualize people's answers

Interview Questions

1. Introducing questions: `Please tell me about when your interest in *X* first began?'; `Have you ever . . .?'; `Why did you do . . .?' .
2. Follow-up/probing questions: getting the interviewee to elaborate his/her answer, such as `Could you say some more about that?'; `What do you mean by that . . .?'; 'Can you give me an example…?'
3. Specifying questions: `What did you do then?'; `How did you open that new Web page?'
4. Direct questions: `Do you find it easy to use the system?'; `Are you happy with the amount of help the documentation provide?'
5. Structuring questions: `I would now like to move on to a different topic'

Interview Questions

6. Indirect questions: `What do most people around here think of that user interface?', perhaps followed up by `Is that the way you feel too?', in order to get at the individual's own view.
7. Interpreting questions: `Do you mean that the way you write an email has changed since you use that system?'; `Is it fair to say that what you are suggesting is that you don't mind switching to this new system but you would not recommend it to other colleagues?'
8. Silence (not a question but important): allow pauses to signal that you want to give the interviewee the opportunity to reflect and amplify an answer.

Quality of an Interviewer

1. Knowledgeable: thoroughly familiar with the focus of the interview (use pilot interviews)
2. Structured: gives purpose for interview; rounds it off; asks whether interviewee has questions.
3. Clear: asks simple, easy, short questions; no jargon.
4. Gentle: lets people finish; gives them time to think; tolerates pauses.
5. Sensitive: listens attentively to what is said and how it is said; is empathetic in dealing with the interviewee.
6. Open: responds to what is important to interviewee and is flexible.

Quality of an Interviewer

7. Remembering: relates what is said to what has previously been said.
8. Interpreting: clarifies and extends meanings of interviewees' statements, but without imposing meaning on them.
9. Balanced: does not talk too much, which may make the interviewee passive, and does not talk too little, which may result in the interviewee feeling he or she is not talking along the right lines.
10. Ethically sensitive: is sensitive to the ethical dimension of interviewing, ensuring the interviewee appreciates what the research is about, its purposes, and that his or her answers will be treated confidentially.

Before and After

► Find:
- a quiet, private space in which to conduct an interview uninterrupted
► Be careful of:
- agreeing to interview someone in their own office
- frequent telephone calls or interruptions
► Spend some time:
- getting hold of a good tape recorder and microphone
- checking the room prior to the interview
- doing a speech recording to test acoustics and carefully positioning the furniture
► Immediately after the interview, note:
- how the interview went
- the setting (busy/quiet, many/few other people in the vicinity, disruptions)
- forgotten topics or good questions for future

Focus Group

► A group discussion/interview to access memories, feelings and perceptions concerning a specific focused topic
► Typically 7-10 people – small enough for everyone to have the opportunity to share insights; large enough to provide diversity of perceptions.
► Moderated by a trained, neutral interviewer (i.e., moderator, facilitator), who asks questions, listens, keeps the conversation on track, and makes sure everyone has a chance to speak
► Must be held in a comfortable, permissive, non-threatening environment
► Three focus groups are considered to be the minimum for a quality study

Ethnographic Observation

► A research process used in the scientific study of human interactions in social settings (ethno = culture, graphy)
► Used extensively in anthropology
► *Purpose* – to describe and explain a facet or segment of group social life
► *Hypotheses and questions* – begin as a broad statement about the purpose of the research, then are allowed to emerge more specifically as data are amassed.
► *Data* - verbal descriptions of people, interactions, settings, objects and phenomena within the context being studied
► *Data Sources* – the people, settings, and relevant objects being observed
► *Data Collection* – done by the researcher through observation combined with interview
► *Data treatment and analysis* – presentation of verbal descriptions and/or logical analysis of information to discover salient patterns and themes

Why Ethnography?

► To define a problem when the problem is not clear
► To define a problem that is complex and embedded in multiple systems or sectors
► To identify participants when the participants, sectors, or stakeholders are not yet known or identified
► To clarify the range of settings where the problem or situation occurs at times when the settings are not fully identified, known, or understood
► To explore the factors associated with the problem in order to understand it
► But
  ▪ Possible bias on the part of the observer (which leads to validity concerns)
  ▪ Generalizability (how generalizable are the findings from a small, purposely selected group)
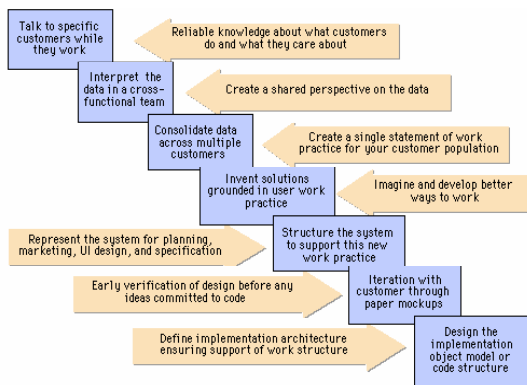
Ethnography Process

1. A question or concern is identified for study
2. A group to study is identified (typically small and purposively selected) → permission is asked
3. The researcher observes the group
   ▪ *Privileged observer* – just observes
   ▪ *Participant observer* – functions as part of the group
4. Researcher watches and listens attentively and records as much detail as possible (this is called *naturalistic observation*). Large amounts of notes are typically generated.
5. Repeat as long as necessary (a week or years).
6. The researcher analyzes the notes, identifies themes, looks for answers to questions, and makes logical inferences.

Contextual Design

Talk to specific customers while they work
Reliable knowledge about what customers do and what they care about
Interpret the data in a cross-functional team
Create a shared perspective on the data
Consolidate data across multiple customers
Create a single statement of work practice for your customer population
Invent solutions grounded in user work practice
Imagine and develop better ways to work
Represent the system for planning, marketing, UI design, and specification
Structure the system to support this new work practice
Early verification of design before any ideas committed to code
Iteration with customer through paper mockups
Define implementation architecture ensuring support of work structure
Design the implementation object model or code structure

Contextual Inquiry principles

► Context: go to the customers' workplace and watch them do their own work.
  ▪ Allows the analyst to experience the rich detail of work rather than the impoverished account of a summary
► Partnership: discuss about their work and engage them in uncovering unarticulated aspects of work.
► Interpretation: develop a shared understanding with user about the aspects of work that matter.
  ▪ Sharing interpretations with users is an important test of our reasoning - and they may fine-tune/correct details
► Focus: direct the inquiry from a clear understanding of your own purpose.
  ▪ Focusing on things that are important/interesting

Contextual inquiry in a nutshell

1. The designer introduces herself, presents the project focus, and asks authorization from the user to record the interview. (15 minutes)
2. The designer explains the methods of Contextual Inquiry. (3 minutes)
3. The user works, the designer observes, and as apprenticeship, makes annotations, diagrams, launches questions and analyzes effects. (one or two hours)
4. The designer confirms her notes with the user, giving him the chance to expand points or conclusions. (10 minutes)

Questionnaires

► Can be administered to large populations
► Paper, email and the web used for dissemination
► Questions can be closed or open
► Closed questions are easier to analyze, and may be done by computer
  ▪ Dichotomous: offering two choices (yes/no)
  ▪ Multiple choice: three or more
  ▪ Likert Scale: balanced units ("scaled")
    ► Odd # allows neutral stance
    ► Even # forces choice → nobody is really neutral
    ► Respondent chooses amount of agreement/ disagreement (usually 5-10 choices)
    ► http://www.socialresearchmethods.net/kb/scallik.php

Questionnaire design

► The impact of a question can be influenced by question order.
► Do you need different versions of the questionnaire for different populations?
► Provide clear instructions on how to complete the questionnaire.
► Strike a balance between using white space and keeping the questionnaire compact.
► Decide on whether phrases will all be positive, all negative or mixed.
► Some examples of HCI questionnaires: http://oldwww.acm.org/perlman/question.html

Encouraging a good response

► Make sure purpose of study is clear
► Explain how anonymity is assured
► Ensure questionnaire is well designed
► Offer a short version for those who do not have time to complete a long questionnaire
► If mailed, include a stamped addressed envelope
► Follow-up with emails, phone calls, letters
► Provide an incentive
► 40% response rate is high, 20% is often acceptable

Online questionnaires

► Advantages
  ▪ Responses are usually received quickly
  ▪ No copying and postage costs
  ▪ Data can be collected in database (automatic formatting)
  ▪ Time required for data analysis is reduced
  ▪ Errors can be corrected easily
► Disadvantages
  ▪ Sampling is problematic if population size is unknown
  ▪ Difficult to prevent individuals from responding more than once
  ▪ Population bias to people w. Internet access