# Prototyping & Evaluating

CZ2004 HCI
Nanyang Technological University

# Prototyping and Evaluating an Interfaces

- Reading - Textbook
  - Chapter 3.1, 3.3, 3.5, 3.9,
  - Chapter 4.1, 4.2, 4.3, 4.4, 4.5, 4.6
- Goal:
  - Familiarize yourself with ethnography, prototyping techniques and evaluation strategies
  - Topics to be overviewed
    - Ethnographic observation
    - Prototyping
    - Evaluating

The Design Process: *Three* Pillars of Design

<image_sentinel_do_not_use id="1"/>
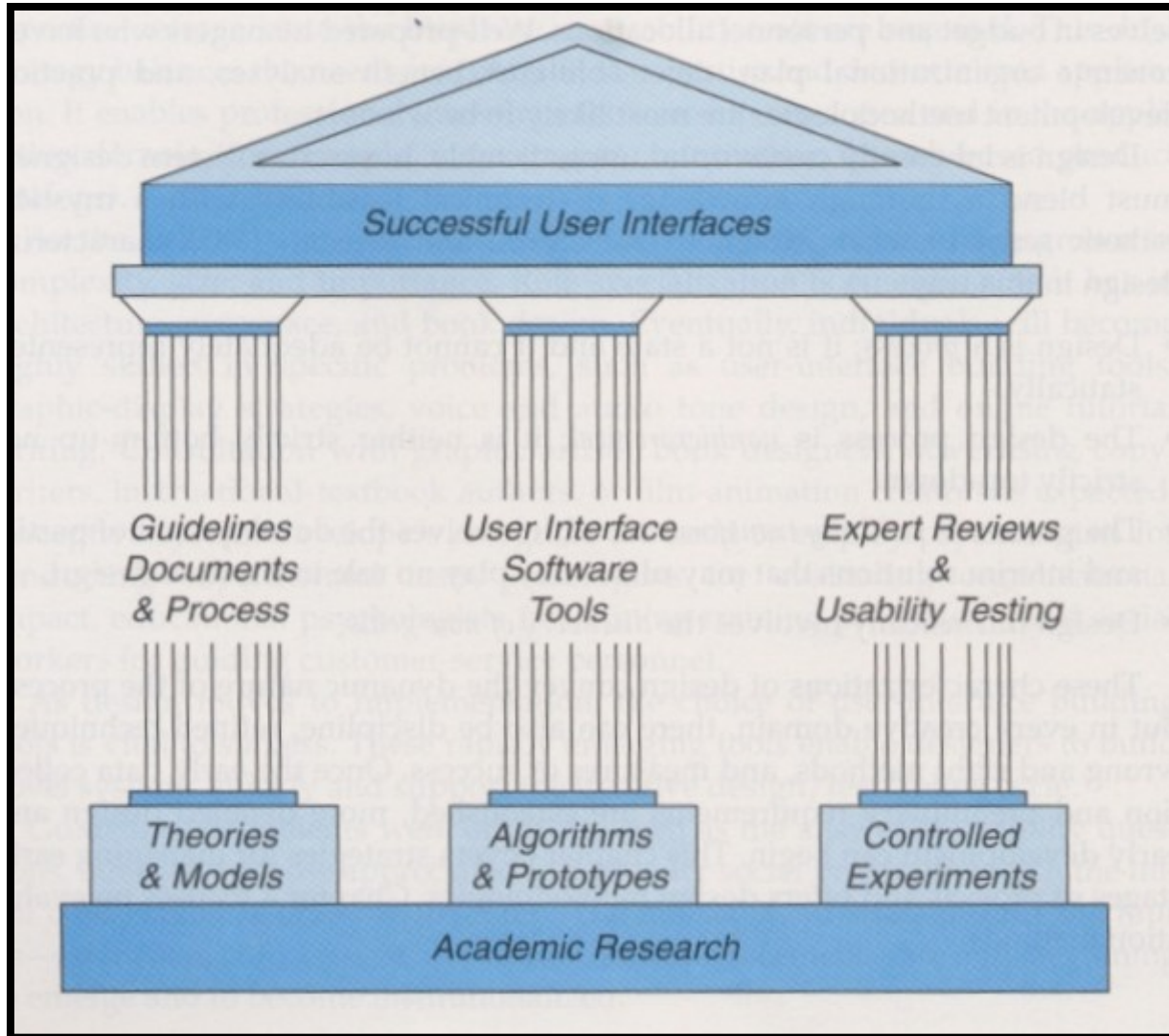
Wait, or, *Four* Pillars?

- HCI is changing, rapidly!
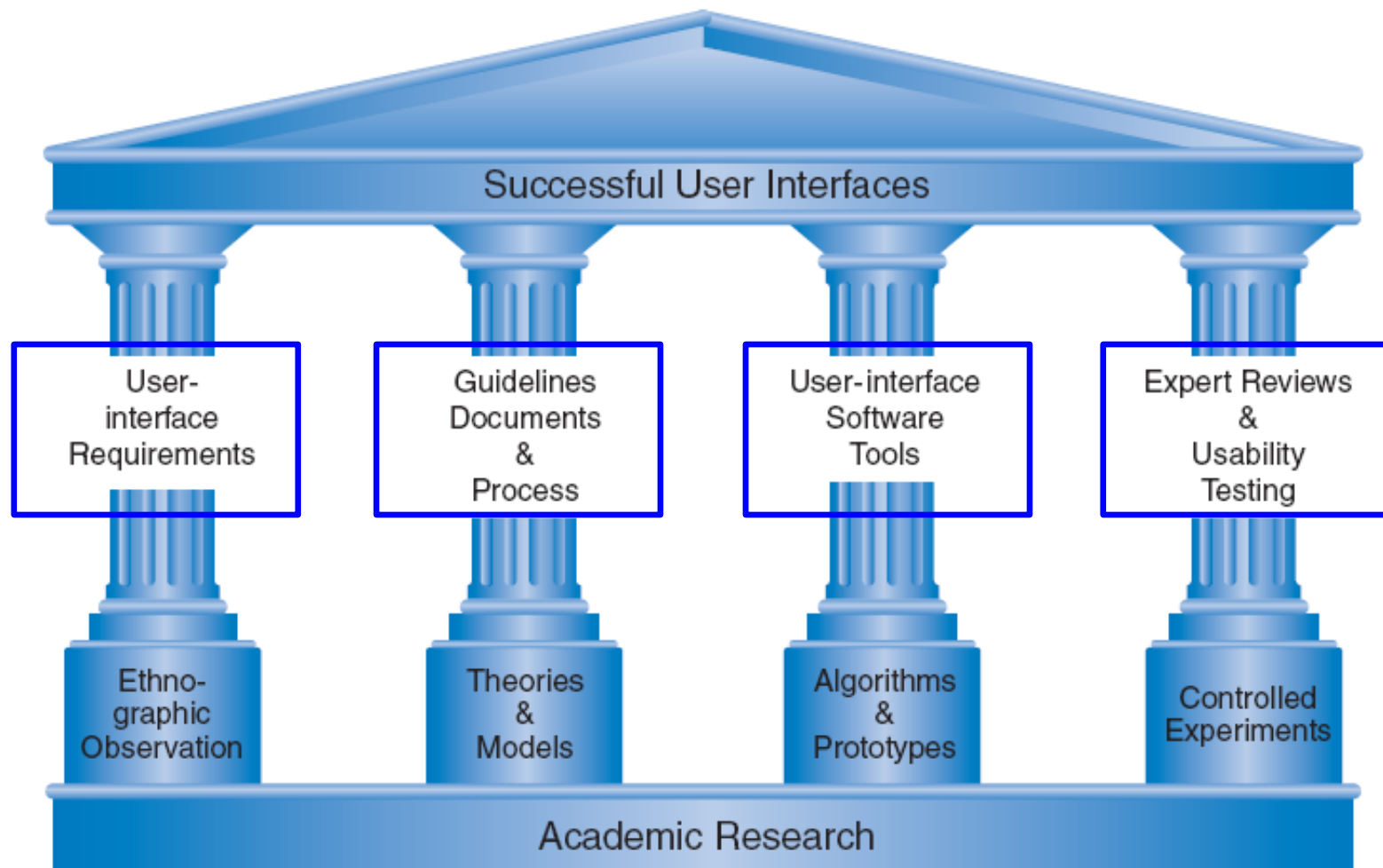
One way to know your users: Ethnographic Observation

- Ethnographic observation steps

- Participatory design

  - Meaning

  - Pros and cons

Meaning of the word

- Ethnography
  - from Greek
  - ἔθνος ethnos = folk/people
  - γράφω grapho = to write
  - A qualitative method aimed to learn and understand cultural phenomena which reflect the knowledge and system of meanings guiding the life of a cultural group
  - Can be *participatory* or *non-participatory*

  - http://en.wikipedia.org/wiki/Ethnography

Participatory Design

Observe "real world" users in their home or office environment

Sends observers to people homes and offices to quietly observe user habits and methods

(General) Ethnographic Observation Steps

- **Preparation**
  - Understand organization policies and work culture.
  - Familiarize yourself with the system and its history.
  - Set initial goals and prepare questions.
  - Gain access and permission to observe/interview.

- **Field Study**
  - Establish rapport with managers and users.
  - Observe/interview users in their workplace and collect subjective/objective quantitative/qualitative data.
  - Follow any leads that emerge from the visits.

Ethnographic Observation Steps (cont.)

- **Analysis**
  - Compile the collected data in numerical, textual, and multimedia databases.
  - Quantify data and compile statistics.
  - Reduce and interpret the data.
  - Refine the goals and the process used.

- **Reporting**
  - Consider multiple audiences and goals.
  - Prepare a report and present the findings

Participatory design: Pros and cons

- more user involvement brings
  - more accurate information about tasks
  - more opportunity for users to influence design decisions
  - a sense of participation that builds users' ego investment in successful implementation
  - potential for increased user acceptance of final system
- on negative side, extensive user involvement may
  - be more costly
  - lengthen the implementation period
  - build antagonism with people not involved or whose suggestions rejected
  - force designers to compromise their design to satisfy incompetent participants

Prototyping

- What is prototyping & Why do we prototype?
- Fidelity
  - Meaning
  - Low-fi
  - Hi-fi
  - Their pros and cons
- Feedback
- *Remember to apply what you learn to labs*

User Interface Software Tools (Pillar II)

PROTOTYPE

- UI Prototype ("mock up")
  - Does **not** have to be functional
  - Simply design the buttons, menus, etc
  - Can well be a hand drawn picture

- Develop prototypes early
  - Very difficult and costly to make changes near the end of a project

- Provides a "vision" for both develops and clients

Why Do We Prototype?

- Get feedback on our design faster
  - saves money
  - Aka, *rapid* development

- Experiment with alternative designs
- Fix problems before code is written
- Keep the design centered on the user

http://www.cc.utah.edu/~asn8200/rapid.html

http://en.wikipedia.org/wiki/Rapid_prototyping

http://www.youtube.com/watch?v=Ps0DSihggio

Fidelity in Prototyping

- Fidelity refers to the level of detail (LoD)

- High fidelity
  - prototypes look like the final product

- Low fidelity
  - artists renditions with many details missing

Fidelity: Low or High?

Problems with Hi-fi Prototypes in the beginning

- Perceptions of the tester/reviewer?
  - formal representation indicates "finished" nature
    - comments on color, fonts, and alignment

- Time?
  - encourage precision
    - specifying details takes more time

- Creativity?
  - lose track of the big picture

Low-fi Storyboards

- What are storyboards?
  - high-level representation of important user actions & consequences

- Where do storyboards come from?
  - Film & animation

- Give you a "script" of important events
  - leave out the details
 - concentrate on the important interactions

Why Use Low-fi Prototypes?

- Traditional methods take too long
  - sketches → prototype (*real code*) → evaluate → iterate

- Can simulate the prototype
  - Sketches (mockup) → evaluate → iterate
  - sketches act as prototypes
    - designer "plays computer"
    - other design team members observe & record

- Kindergarten implementation skills
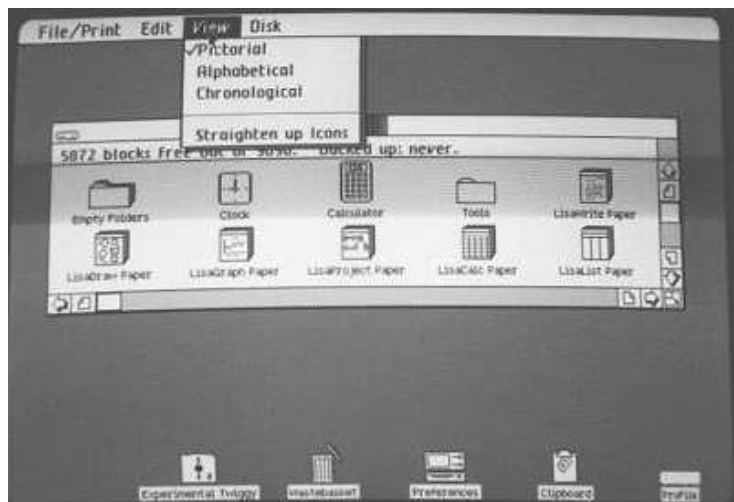 - allows *non-programmers* to participate

High-Fidelity Mockup

- After interface is established, still room for more mockup

- High-Fidelity can look much more similar to final product

- Paper-based may still be used OK, but most likely using real code and software

Mac (LISA) Prototypes
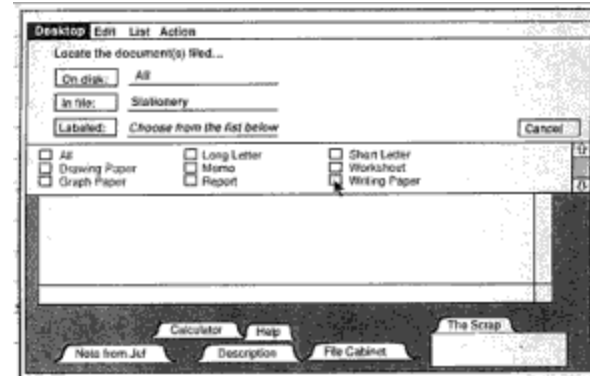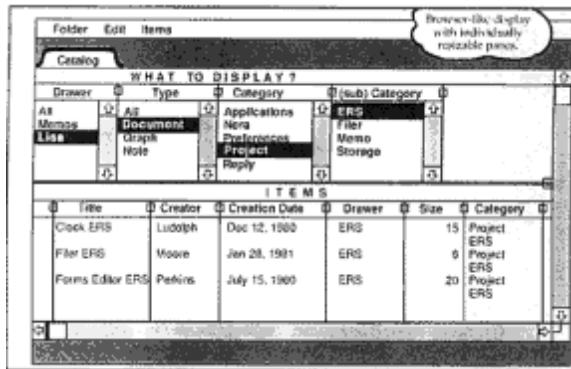
Paper-based high-fidelty "mock ups" from 1980 and 1981 Apple OS

Lisa release 1983

OSX (2000)

Prototypes Feedback

- Previews and feedback from prototypes is crucial
  - Even low fidelity mockups
  - If possible, get feedback from experts & average users

- Large websites carry out "pilot" launches to get feedback and find bugs

- Google "Gmail" is a great example
  - Request feedback and requests/suggestions for improvements
  - Launched April 1, 2004, throw away "beta" on July 7, 2009
  - 260 million users (Oct. 2011)

Evaluating

- Expert Review
- User Labs
- Questionnaires/Surveys
- Acceptance Testing
- Continuous feedback (evaluation during active use)

Evaluation: Introduction

- Designers can become so entranced with their creations that they may fail to evaluate them adequately.
- Experienced designers have attained the wisdom and humility to know that extensive testing is a necessity.
- The determinants of the evaluation plan include:
  - stage of design (early, middle, late)
  - novelty of project (well defined vs. exploratory)
  - number of expected users
  - criticality of the interface (life-critical medical system vs. museum exhibit support)
  - costs of product and finances allocated for testing
  - time available
  - experience of the design and evaluation team

Evaluation: Introduction (cont.)

- The range of evaluation plans might be from an ambitious two-year test to a few days test.

- The range of costs might be from 20% of a project down to 5%.
  - Budget for user evaluation!

- Is it worth it? Well, remember this . . . good design pays off.

Evaluation: Introduction (cont.)

- Troubling aspect of testing: uncertainty remains even after exhaustive testing by multiple methods.
  - Perfection is not possible in complex human endeavors, so must continue assessing and repairing problems during lifecycle of interface
  - Decision must be made about completing prototype testing and delivering the product, even though problems may continue to be found
  - Most testing methods will account fro normal usage, but performance in unpredictable situations with high levels of input such as nuclear reactor control, is extremely hard to test.

Expert Reviews

- Expert reviews entail one-half day to one week effort, although a lengthy training period may sometimes be required to explain the task domain or operational procedures
- There are a variety of expert review methods to chose from:
  - **Heuristic evaluation**
    - Experts reviewers personal critic
  - **Guidelines review**
    - Make sure UI adheres to established guidelines
  - **Consistency inspection**
    - Check for consistency through-out interface
  - **Cognitive walkthrough**
    - Simulate performing certain tasks
  - **Formal usability inspection**
    - UI designers defend their choices against a "hostile" expert

1. Heuristic evaluation

- Expert reviewer gives personal criticism
  - Could be expert designer – like Alan Cooper
  - Could be expert "user" (i.e. doctor for medical product)
  - Often have more than one reviewer
- The expert reviewer spends time evaluating your interface
  - Gives feedback, overall impression, concerns, maybe according to 'eight golden rule'
  - Remember this is an expert, so you are paying to hear what they say
- "heuristic" evaluation
  - No true formal approach – thus the term "heuristic"
  - But effective
  - This is often called "Discount Evaluation" because it is typically cheaper than a full user study

2. Guideline Review

- Assume you have a guideline document

- Scrutinize the UI to make sure it adheres to the guidelines

- May require a "bird's eye" view of the UI
  - This is where all possible windows of the UI are printed out, laid out on the floor or pinned to walls.

- *The guideline may be of a thousand items, so it may take expert reviewers sometime to master the guideline, and days or weeks to review a large interface.*

3. Consistency Checking

- Similar to "guideline review", but check for consistency in the UI

- Consistency involves:
  - Terminology, icons, color scheme, layout, input/output format, etc
  - Within the interface as well as in the training materials (e.g. tutorial) and online help
  - Software tools may help automate the process

4. Cognitive Walkthrough

- Ask expert reviewer to perform certain tasks, stimulating users walking through the interface

- Watch how they do it

- See if they behave they way you thought they would!
  - Get verbal feedback
  - User talks out loud the whole time explaining his/her thought process

- May see usage patterns you didn't expect!

5. Formal usability inspection

- Courtroom like setting

- Expert reviewers ask questions to the designers to justify their design decisions in an adversarial manner.

- A lengthy process, but can be educational to inexperience designers.

Expert Reviews – issues

- Expert reviews can be scheduled at several points in the development process when experts are available and when the design team is ready for feedback.
- Different experts tend to find different problems in an interface, so *3-5* expert reviewers can be highly productive, as can complementary usability testing.
- The dangers with expert reviews are that the experts may not have an adequate understanding of the task domain or user communities.
  - Coming in many flavors, experts may give conflicting advice
  - "For every Ph.D., there is an equal and opposite Ph.D."
- Even experienced expert reviewers have great difficulty knowing how typical users, especially first-time users will really behave.

Usability Testing and Laboratories

- The emergence of usability testing and laboratories since the early 1980s.
- Usability testing not only sped up many projects but has also produced dramatic cost savings.
  - Traditional managers and developers resisted at first, saying the usability testing may take time and resource away from development
  - They changed their mind when experience grew and successful projects gave credit to the testing process.
- The movement towards usability testing stimulated the construction of usability laboratories.

Usability lab

- A typical modest usability lab would have two 10 by 10 foot areas, one for the participants to do their work and another, separated by a half-silvered mirror, for the testers and observers

- Participants should be chosen to represent the intended user communities, with attention to
  - background in computing, experience with the task, motivation, education, and ability with the natural language used in the interface.

A case study: Make users work for you

- Google, in 2005, has taken its first foray into the physical world with the launch of an Internet cafe-style computing booth in London's Heathrow Airport.

- Although the search specialist claims the project is primarily about helping travelers use their time more productively at the airport, it admitted that Google Space will also act as a physical testing lab for its new applications.

- "Google Space will help people make wasted time more useful. And for Google, Space is a live lab where people can test our most-up-to-date products, and give us their feedback, "said Lorraine Twohill, Google's European director of marketing.
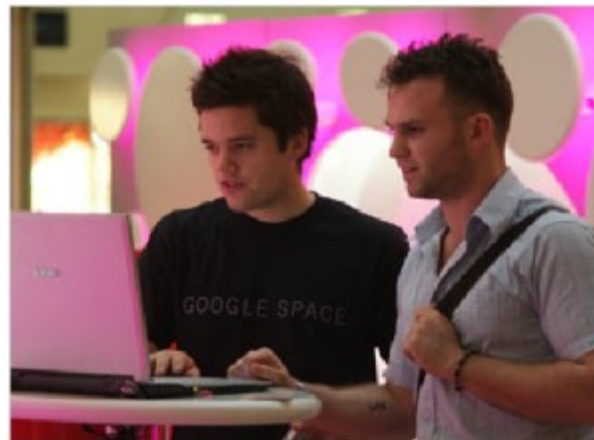
Issues in usability testing and labs

- Participation should always be voluntary, and informed consent should be obtained.
- Professional practice is to ask all subjects to read and sign a statement like this one:
  - I have freely volunteered to participate in this experiment.
  - I have been informed in advance what my task(s) will be and what procedures will be followed.
  - I have been given the opportunity to ask questions, and have had my questions answered to my satisfaction.
  - I am aware that I have the right to withdraw consent and to discontinue participation at any time, without prejudice to my future treatment.
  - My signature below may be taken as affirmation of all the above statements; it was given prior to my participation in this study.

Usability lab variations: videotaping

- Videotaping participants performing tasks is often valuable for later review and for showing designers or managers the problems that users encounter.
  - Reviewing videotaping is tedious, so careful logging and annotation is vital on finding critical incidents
    - Tools for automatic time stamping activities (typing, mousing, reading manuals, etc.)
  - Participants may be anxious about the video camera at the start of the test…minutes later, focusing on the tasks

Usability lab variations: Think Aloud

- Think Aloud: participants carry out tasks while saying what they are thinking, and tester records thoughts
  - Tester does not take over or give instructions, but prompt and listen for clues how they are dealing with the interface.
    - "tell me what you are thinking"
    - "…so this is?"
  - Think-aloud technique yield interesting clues for the tester
    - *For example, "This web page text is too small…so I am looking for something on the menus to make the text bigger…maybe it's on the top in the icons…I can't find it…so I'll just carry on".*

Usability lab variations: more

- Variant forms of usability testing:
  - Paper mockups
    - Early usability study; inexpensive, rapid.
    - Flipping the (mockup of) screen displays to get reactions to wording, layout, etc.
  - Discount usability testing
    - Only 3 to 6 test participants
  - Competitive usability testing
    - Comparing new interface to previous versions or to similar products
  - Universal usability testing
    - Diverse users, hardware/software platform, networks, etc.
  - Field test and portable labs
    - New interface to work in realistic environments for a fixed trial period.
    - Test of new software or consumer products
  - Remote usability testing
    - Tests online; less control over user behavior and observation of reaction
  - Can-you-break-this tests
    - Game design: challenge energetic teenagers to beat new games, finding fatal flaws

Survey Instruments

- Written user surveys are a familiar, inexpensive and generally acceptable companion for usability tests and expert reviews.

- Keys to successful surveys
  - Clear goals in advance
  - Development of focused items that help attain the goals.

- Users could be asked for their subjective impressions about specific aspects of the interface such as the representation of:
  - task domain objects and actions
  - syntax of inputs and design of displays.

Survey Instruments (cont.)

- Other goals would be to ascertain
  - users background (age, gender, origins, education, income)
  - experience with computers (specific applications or software packages, length of time, depth of knowledge)
  - job responsibilities (decision-making influence, managerial roles, motivation)
  - personality style (introvert vs. extrovert, risk taking vs. risk aversive, early vs. late adopter, systematic vs. opportunistic)
  - reasons for not using an interface (inadequate services, too complex, too slow)
  - familiarity with features (printing, macros, shortcuts, tutorials)
  - their feeling state after using an interface (confused vs. clear, frustrated vs. in-control, bored vs. excited).

Surveys (cont.)

- Online surveys avoid the cost of printing and the extra effort needed for distribution and collection of paper forms.

- Many people prefer to answer a brief survey displayed on a screen, instead of filling in and returning a printed form,

  - although there is a potential bias in the sample.

Survey Example

10. Please rank your use and usability of each page: (1=Minimal/Difficult, 5=Extensive/Easy)

	Use	Usability
Home page	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
About the Coalition	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
Member governors	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
Representatives	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
Publications	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
News Releases	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
Letters/Speeches	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
Ethanol Information	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
State/Federal Legislation	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
Comments	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
Members Only	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
Links	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5
Search	○1 ○2 ○3 ○4 ○5	○1 ○2 ○3 ○4 ○5

11. Please rate the site on these features:

Layout and design	○ Poor	○ Fair	○ Good	○ Excellent	○ Don't Know
Ease of finding information	○ Poor	○ Fair	○ Good	○ Excellent	○ Don't Know
Usefulness of information	○ Poor	○ Fair	○ Good	○ Excellent	○ Don't Know
Login feature	○ Poor	○ Fair	○ Good	○ Excellent	○ Don't Know
Getting assistance with the website	○ Poor	○ Fair	○ Good	○ Excellent	○ Don't Know
Timeliness of information	○ Poor	○ Fair	○ Good	○ Excellent	○ Don't Know
Navigation	○ Poor	○ Fair	○ Good	○ Excellent	○ Don't Know

Acceptance Test

- For large implementation projects, the customer or manager usually sets objective and measurable goals for hardware and software performance.

- If the completed product fails to meet these acceptance criteria, the system must be reworked until success is demonstrated.

- Rather than the vague and misleading criterion of "user friendly," measurable criteria for the user interface can be established for the following:
  - Time to learn specific functions
  - Speed of task performance
  - Rate of errors by users
  - Human retention of commands over time
  - Subjective user satisfaction

  - Well, do you remember the above 5 items?

Acceptance Test (cont.)

- Example:
  "Ten participants will be recalled after one week, and asked to carry out a new set of benchmark tasks. In 20 minutes, at least 8 of the participants should be able to complete the tasks correctly."
- In a large system, there may be 8 or 10 such tests to carry out on different components of the interface and with different user communities.
- Once acceptance testing has been successful, there may be a period of field testing before national or international distribution.

Evaluation During Active Use

- Successful active use requires constant attention from dedicated managers, user-services personnel, and maintenance staff.
- Perfection is not attainable, but percentage improvements are possible.

Evaluation During Active Use: strategy

- Interviews and focus group discussions
  - Interviews with individual users can be productive because the interviewer can pursue specific issues of concern.
  - Group discussions are valuable to ascertain the universality of comments.
- Continuous user-performance data logging
  - The software architecture should make it easy for system managers to collect data about
    - The patterns of system usage
    - Speed of user performance
    - Rate of errors
    - Frequency of request for online assistance
  - A major benefit is guidance to system maintainers in optimizing performance and reducing costs for all participants.
- Online or telephone consultants
  - Many users feel reassured if they know there is a human assistance available
  - On some network systems, the consultants can monitor the user's computer and see the same displays that the user sees

Strategy (cont.)

- Online suggestion box or e-mail trouble reporting
  - Electronic mail to the maintainers or designers.
  - For some users, writing a letter may be seen as requiring too much effort.
- Discussion group and newsgroup
  - Permit postings of open messages and questions
  - Some are independent, e.g. America Online and Yahoo!
  - Topic list
  - Sometimes moderators
  - Social systems
  - Comments and suggestions should be encouraged.

Summary

- Four pillars
  - How do they form a design process
  - How do they relate to our lecture modules
- Ethnographic observation
- Prototyping
  - Importance and Significance
  - Fidelity
  - Feedback
- Evaluating
  - Importance and Significance
  - Several different approaches outlined
    - Expert Review
    - User Labs
    - Questionnaires/Surveys
    - Acceptance Testing
    - Continuous feedback (evaluation during active use)