

CHAPTER 5:

Evaluation and the User Experience

*Designing the User Interface:
Strategies for Effective Human-Computer Interaction*

Sixth Edition

Ben Shneiderman, Catherine Plaisant,
Maxine S. Cohen, Steven M. Jacobs, and Niklas Elmqvist
*in collaboration with
Nicholas Diakopoulos*

Addison Wesley
is an imprint of



© 2017 Pearson Education, Inc., Hoboken, NJ. All rights reserved.

Evaluation and the User Experience

Topics

1. Introduction
2. Expert Reviews and Heuristics
3. Usability Testing and Laboratories
4. Survey Instruments
5. Acceptance Tests
6. Evaluation During Active Use and Beyond
7. Controlled Psychologically-Oriented Experiments

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

Introduction (concluded)

- Usability evaluators must broaden their methods and be open to non-empirical methods, such as user sketches, consideration of design alternatives, and ethnographic studies.
 - Recommendations needs to be based on observational findings
- The design team needs to be involved with research on the current system design drawbacks
 - Tools and techniques are evolving
 - The range of evaluation plans might be anywhere from an ambitious two-year test with multiple phases for a new national air-traffic control system to a three-day test with six users for a small internal web site
 - The range of costs might be from 20% of a project down to 5%.
- Usability testing has become an established and accepted part of the design process

Expert Reviews and Heuristics

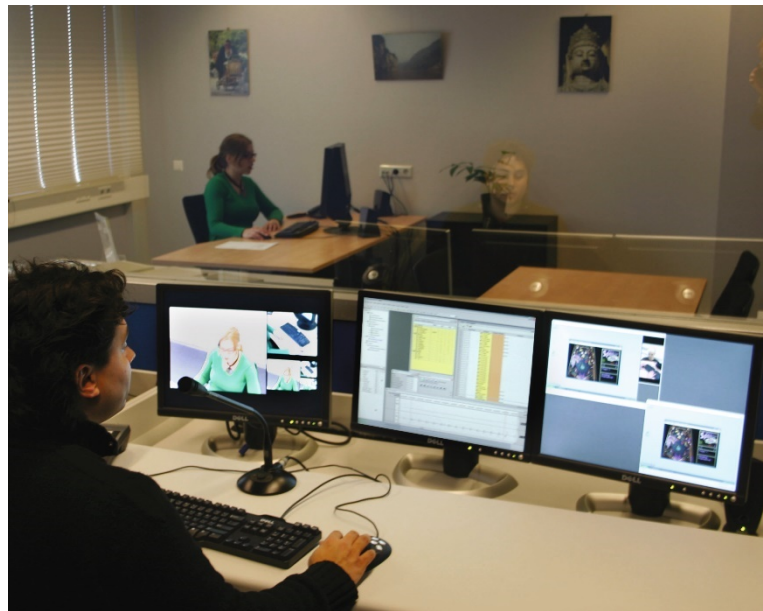
- While informal demos to colleagues or customers can provide some useful feedback, more formal expert reviews have proven to be effective
- Expert reviews entail one-half day to one week effort, although a lengthy training period may sometimes be required to explain the task domain or operational procedures
- There are a variety of expert review methods to choose from:
 - Heuristic evaluation
 - Guidelines review
 - Consistency inspection
 - Cognitive walkthrough
 - Formal usability inspection

Expert Reviews and Heuristics (concluded)

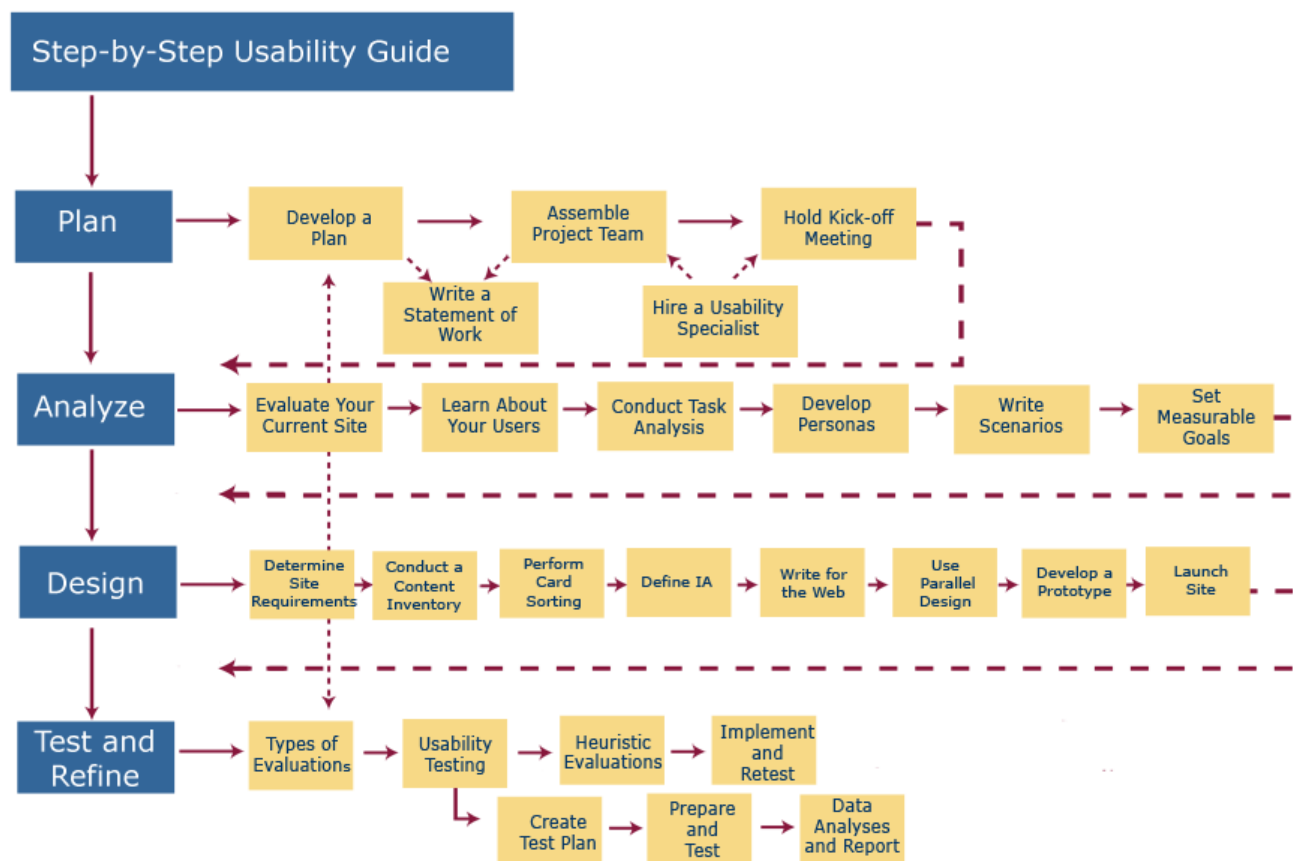
- Expert reviews can be scheduled at several points in the development process when experts are available and when the design team is ready for feedback
- Different experts tend to find different problems in an interface, so 3-5 expert reviewers can be highly productive, as can complementary usability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

Usability Testing and Laboratories

- The usability lab consists of two areas: the testing room and the observation room
 - The testing room is typically smaller and accommodates a small number of people
 - The observation room, can see into the testing room typically via a one-way mirror. The observation room is larger and can hold the usability testing facilitators with ample room to bring in others, such as the developers of the product being tested



Step-by-Step Usability Guide from <http://usability.gov/>



Usability Testing and Laboratories (continued)

- This shows a picture of glasses worn for eye-tracking
 - This particular device tracks the participant's eye movements when using a mobile device
 - Tobii is one of several manufacturers



Usability Testing and Laboratories (continued)

- Eye-tracking software is attached to the airline check-in kiosk
 - It allows the designer to collect data observing how the user “looks” at the screen
 - This helps determine if various interface elements (e.g. buttons) are difficult (or easy) to find



Usability Testing and Laboratories (continued)

- The special mobile camera to track and record activities on a mobile device
 - Note the camera is up and out of the way still allowing the user to use their normal finger gestures to operate the device



Usability Testing and Laboratories (continued)

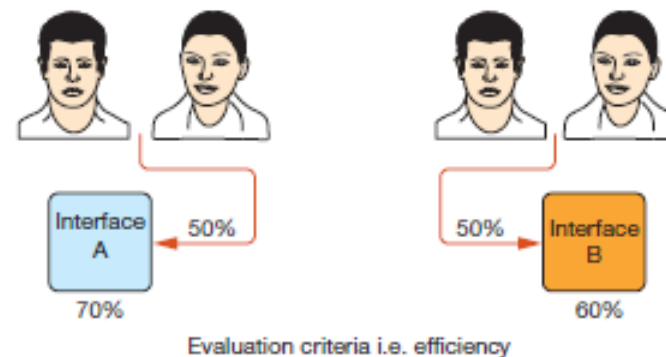
- The emergence of usability testing and laboratories since the early 1980s
- Usability testing not only sped up many projects but that it produced dramatic cost savings
- The movement towards usability testing stimulated the construction of usability laboratories
- A typical modest usability lab would 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 and experience with the task
 - motivation, education, and ability with the natural language used in the interface.

Usability Testing and Laboratories (continued)

- Participation should always be voluntary, and informed consent should be obtained
- Professional ethics practice is to ask all subjects to read and sign a statement like this:
 - 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.
- Institutional Review Boards (IRB) often governs human subject test process

Usability Testing and Laboratories (concluded)

- Videotaping participants performing tasks is often valuable for later review and for showing designers or managers the problems that users encounter
 - Use caution in order to not interfere with participants
 - Invite users to think aloud (sometimes referred to as concurrent think aloud) about what they are doing as they are performing the task
- Many variant forms of usability testing have been tried:
 - Paper mockups
 - Discount usability testing
 - Competitive usability testing
 - A/B testing
 - Universal usability testing
 - Field test and portable labs
 - Remote usability testing
 - Can-you-break-this tests
 - Think-aloud and related techniques
- Usability test reports



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

Examples of the specific satisfaction scale questions:													
5.4	Messages which appear on display:	confusing	clear	1	2	3	4	5	6	7	8	9	NA
5.4.1	Instructions for commands or choice:	confusing	clear	1	2	3	4	5	6	7	8	9	NA

Survey Instruments (continued)

- Other goals would be to ascertain:
 - User 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 or extrovert, risk taking or risk averse, early or late adopter, systematic or opportunistic)
 - Reasons for not using an interface (inadequate services, too complex, too slow)
 - Familiarity with features (printing, macros, shortcuts, tutorials)
 - Feeling state after using an interface (confused or clear, frustrated or in-control, bored or excited)

Survey Instruments (concluded)

- 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
- A survey example is the Questionnaire for User Interaction Satisfaction (QUIS)
 - <http://lap.umd.edu/quis/>
- There are others, e.g. Mobile Phone Usability Questionnaire (MPUQ)

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

Acceptance Test (concluded)

- 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 and Beyond

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

Evaluation During Active Use and Beyond (continued)

- 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 chat consultants, e-mail, and online suggestion boxes
 - 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

Evaluation During Active Use and Beyond (continued)

- 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 groups, wikis and newsgroups
 - 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

Evaluation During Active Use and Beyond (concluded)

- Example output of an automated evaluation tool from TechSmith's Morae
 - The item being measured is mouse clicks.
 - This shows the view for task 2 (selected in the tabbed bar). Obviously, the other 3 tasks could also be displayed. These are the values for participant 4.
 - The drop down list box would allow the evaluator to choose the mouse clicks for other participants.
 - Across the horizontal axis time is shown



Controlled Psychologically-oriented Experiments

- Scientific and engineering progress is often stimulated by improved techniques for precise measurement
- Rapid progress in the designs of interfaces will be stimulated as researchers and practitioners evolve suitable human-performance measures and techniques

Controlled Psychologically-oriented Experiments (continued)

- The outline of the scientific method as applied to human-computer interaction might comprise these tasks:
 - Deal with a practical problem and consider the theoretical framework
 - State a lucid and testable hypothesis
 - Identify a small number of independent variables that are to be manipulated
 - Carefully choose the dependent variables that will be measured
 - Judiciously select subjects and carefully or randomly assign subjects to groups
 - Control for biasing factors (non-representative sample of subjects or selection of tasks, inconsistent testing procedures)
 - Apply statistical methods to data analysis
 - Resolve the practical problem, refine the theory, and give advice to future researchers

Controlled Psychologically-oriented Experiments (concluded)

- Controlled experiments can help fine-tuning the human-computer interface of actively used systems
- Performance could be compared with the control group
- Dependent measures could include performance times, user-subjective satisfaction, error rates, and user retention over time