

**CSC 484 Lecture Notes Week 8**  
**The DECIDE Evaluation Framework;**  
**Usability Testing and Field Studies;**  
**Recap of Analytic Evaluation**

I. **Relevant reading** -- Chapters 13-15 of the textbook.

II. **Introduction to Chapter 13 (Section 13.1).**

- A. This chapter is largely a recapitulation of preceding chapters.
- B. It provides a high-level organizational framework for evaluation, called "DECIDE".
- C. In presenting DECIDE, the authors provide a few new pieces of information, not explicitly mentioned in preceding chapters.
- D. You may find the DECIDE framework helpful in organizing the evaluation part of your final project report.

III. **Definition of DECIDE (Section 13.2).**

- A. Again, the purpose of the framework is to provide some high-level organizational guidance for evaluation of interaction designs.
- B. DECIDE has six steps:
  1. **D**etermine the *goals*.
  2. **E**xplore the *questions*.
  3. **C**hoose the *evaluation approach* and *methods*.
  4. **I**dentify the *practical issues*.
  5. **D**ecide how to deal with the *ethical issues*.
  6. **E**valuate, analyze, interpret and present the *data*.

IV. **Determine the goals (Section 13.2.1).**

- A. We have discussed this amply, particularly in last weeks notes on chapters 7 and 8 of the book.

V. **Explore the questions (Section 13.2.2).**

- A. We also discussed this last week, particularly in the context of properly framing questions on a questionnaire.
- B. An important point to reiterate is that of not forgetting to ask *fundamental* questions of the subjects in a usability study.
  1. In the heat of project development, a team may get so fully immersed in the work as to lose sight of basic questions to ask.
  2. For example, "Would you use this product?", "If so, how often?".
- C. And of course such questions should be asked on questionnaire in an analyzable form, as in
  - "I would use this product for its intended purpose.  Strongly disagree ...  Strongly agree"
  - "I would use this product  daily  weekly  monthly  occasionally"

VI. **Choosing the appropriate methods (Section 13.2.3).**

- A. Been here, done this.

**VII. Identify the practical issues (Section 13.2.4).**

- A. The major practical issue for 484 -- DO DRESS REHEARSAL of your usability study.
  - 1. I.e., each team member act independently as a study participant.
  - 2. If possible, enlist the help of people you know outside the class, by whatever means of cajolation you have at your disposal.
- B. **Practical Users issues.**
  - 1. Page 631 notes the following bits of information about task length in usability studies.
    - a. 10 minutes is too short, 2 hours too long.
    - b. This means the 50-minute time slots we have for 484 studies are just about right.
  - 2. Page 631 also recounts the dilemma of how to study people's behavior without influencing it.
  - 3. The lesson in particular for 484 studies is leave study participants alone as much as possible, while they're performing the study tasks.
- C. **Practical issues of facilities and scheduling.**
  - 1. Plan the logistics of your usability study thoroughly.
  - 2. Think through the room layout and how participants will be placed.
  - 3. Plan all equipment placement.
  - 4. Assign study monitoring duties to team members.
  - 5. Determine how the questionnaires will be administered and collected.
- D. **Practical issues of expertise.**
  - 1. For 484 study data analysis, take advantage of Heather Smith's statistical expertise.
  - 2. She has regular open office hours, and is available by appointment.
  - 3. By all accounts, her advice is professional and very helpful.

**VIII. Decide how to deal with the ethical issues (Section 13.2.5).**

- A. You've had a class in this; apply what you learned in 300 to your work in 484.
- B. Activity 13.6 describes practice that you **should follow** in your 484 studies, particularly the 2d3d team:
  - 1. Assign each study participant a code number.
  - 2. Have them put their number, not name, on the questionnaire, and any other data you collect from them.
  - 3. Keep the name-to-code correlation information separate from the collected data.
- C. As noted in the Milestone 3 writeup, you are required to have controlled-study participants sign an informed consent form.
  - 1. For your fellow 484 students, this is an academic exercise, but worth doing explicitly to focus your attention on its necessity in real studies.
  - 2. It's in fact necessary for the 2d3d study, since you're using study participants from outside of class.
  - 3. A consent form is **not necessary** for field-study interviews, in particular those done by the swat team.
- D. There is a summary of ethical points to consider on Pages 637-638 of the book:
  - 1. Tell participants the study goals, how long it will take, how the data will be reported and analyzed; offer them a copy of the final report (most likely in the form of a webpage link).
  - 2. Explain that any revealed or discovered personal information will be kept confidential.
  - 3. Make sure participants know they are free to stop at any time.
  - 4. Consider the appropriateness of any offered incentives.
  - 5. Do not report quotes that could inadvertently reveal the identity of participants; use numbers or fictitious names in any explicit quotes.

6. Always ask permission to quote a participant; offer a pre-release copy of the report to those quoted, so they can check that they were quoted correctly.

#### IX. Evaluate, interpret, and present the data (Section 13.2.6).

- A. Your 484 studies are simple academic exercises, most likely not subject to scrutiny by the outside world.
- B. However, it is worth considering the following criteria for data evaluation and interpretation.
  1. **Reliability** -- bottom line, is it reproducible by others?
  2. **Validity** -- did you prove what you set out to prove, i.e., did you validate any claims that you made at the outset of the study?
  3. **Biases** -- did you do your utmost not to let biases corrupt the study, the data gathering, its analysis, and interpretation?
  4. **Scope** -- how much can your findings be generalized, if at all?
  5. **Ecological validity** -- does the environment in which the study is conducted effect the results, e.g., do you get better results if you fed the subjects well?
- C. Regarding the last point on "ecological validity", the Wikipedia article on the *Hawthorne effect* is one of the more cogent articles I've read at that venue; I recommend it.

#### X. Introduction to Chapter 14, on Usability Testing and Field Studies (Section 14.1)

- A. The primary focus of this chapter is on usability studies of finished products.
- B. As such, many of the specifics do not apply directly to the 484 studies that are based on prototypes of *proposed* products.
- C. Nevertheless, there is some useful information for 484 teams who will be conducting studies that involve some form of quantitative analysis, particularly 2d3d mobility, and touchten.

#### XI. Usability testing (Section 14.2).

- A. As presented in previous notes and chapters, the key components are
  1. user tests
  2. satisfaction questionnaires
  3. interviews
- B. For fully quantifiable user tests, the following types of data are gathered:
  1. time to complete a task
  2. time to complete, after being away from the product for a specified time
  3. number of errors per task
  4. number or errors per unit of time
  5. number of navigations to online help or manuals
  6. number of users making a particular error
  7. number of users completing a task successfully
- C. The number of participants in usability studies varies.
  1. For the type of quantitative testing described in Chapter 14, the book cites a 1999 book by Dumas and Redish that considers 5-12 users acceptable.
  2. On his web page, Nielsen basically concurs with this, saying 5-15 users is sufficient (see "*Why You Only Need to Test With 5 Users*" at [www.useit.com/alertbox/20000319.html](http://www.useit.com/alertbox/20000319.html)).
  3. Both these sources are talking about usability studies that focus on specific features, using a number of small tests.

4. For studies involving statistical analysis, the number of participating subjects depends on the type of analyses to be performed, and the degree of statistical certainty to be achieved.
  - a. Generally, statistical analyses require a sample size of more than 15 subjects.
  - b. There are well-known formulae for calculating experimental sample sizes.
  - c. Russ Length's web page has a Java applet, and links to a number of other related sites; see [www.stat.uiowa.edu/~rlenth/Power](http://www.stat.uiowa.edu/~rlenth/Power)
- D. The venues of usability studies vary widely.
  1. Large companies, like Microsoft, have large dedicated spaces, equipped to the hilt with recording equipment, sound proofing, and any other conceivable experimental amenity.
  2. On the other end of the spectrum is the "lab-in-a-suitcase" approach, where studiers travel to users' sites to conduct their work.
  3. There is also the remote monitoring approach, where study participants work in their own environment, with those conducting the study monitoring and collecting data remotely.

## XII. Usability testing of a large website (Section 14.2.1).

- A. The book walks through a concrete example, of interest, but not direct relevance to your 484 studies.
- B. The structure of the presentation is a good review of the steps involved in a usability study, as have been outlined in chapters 7, 8, and 13:
  1. *Establishing goals and questions*
  2. *Selection of participants*
  3. *Development of the tasks*
  4. *The test procedure*
  5. *Data collection*

## XIII. Conducting experiments in usability testing (Section 14.2.2).

- A. A usability study is sometimes carried out in the form of a scientific experiment.
- B. This involves testing a specific hypothesis.
- C. A basic hypothesis is stated in terms of two variables.
- D. E.g., "*Reading text displayed in 12-point Helvetica font is faster than Reading text displayed in 12-point Times New Roman.*"
- E. The variables are classified as *dependent* or *independent*.
  1. The value of an independent variable is *selected* in advance by the experimenter, e.g., the font type in the preceding example.
  2. The value of a dependent variable is *measured* in the experiment, e.g., the time taken to read the text.
- F. An hypothesis can be stated in *null* and *alternative* forms.
  1. The null hypothesis states the opposite of an experimenter's proposed, a.k.a., alternative hypothesis.
  2. In the preceding example, the null hypothesis is that there is no difference between reading times with the two fonts.
  3. Statistically, the null hypothesis provides a baseline for measuring the significance of observed findings.
  4. I.e., significance is defined in terms of how often observed data support the null hypothesis.
  5. If this happens rarely, it allows the experimenter to invoke a form of proof- by-contradiction.
  6. I.e., if gathered evidence can be shown rarely to support the null hypothesis, then the alternative hypothesis is assumed true, within the determined degree of statistical significance.

- G. HCI experiments very often involve multiple variables.
  - 1. There can be more than one dependent variable, and more than one independent variable.
  - 2. In HCI experiments, there are very typically other unmeasured variables in an experiment, the values of which must be kept constant.
  - 3. E.g., font color and screen resolution are such variables in the preceding example.
- H. The significant challenges of experimental design are to
  - 1. identify all the variables that may effect the experimental outcome,
  - 2. set up the experimental conditions so that the values of all unmeasured variables remained fixed.
- I. The book provides further details and examples of the statistical design of some simple HCI experiments.
- J. Several of the 484 research readings have larger-scale examples of statistical experimental design.
- K. And there are many web and textbook resources on the subject.

#### XIV. Field studies (Section 14.3).

- A. This section is largely a recap of material presented in preceding chapters.
- B. Here are some important points to remember about field studies, most notably for the sultans of swat:
  - 1. Tell participants what they will be asked to do, and about how long it will take.
  - 2. Have a plan of the questions you want to ask, but be prepared to be flexible as circumstances dictate.
  - 3. Let the participants "do their own thing" during the specific period of prototype usage.
  - 4. Observe participants as unobtrusively a possible, but *do* observe in some form.
  - 5. Record the sessions with notes, and other forms of recording that the participants consider acceptable.
- C. The larger-scale examples in Section 14.3 are not directly applicable to 484.
- D. Neither are the theoretical frameworks used -- *activity theory* and *semiotic engineering*, but worth a read.

#### XV. Introduction to Chapter 15, on Analytic Evaluation (Section 15.1).

- A. This topic was the subject 484 Assignment 1, for which you already read Section 15.2.
- B. The gist of analytic evaluation is that it does not involve *actual* end users, where "actual" means real people who will use a product for its intended purpose.
- C. Rather, analytic evaluation is conducted by assumed product and domain experts, using inspections or models to evaluate a product.

#### XVI. Heuristic Evaluation (Section 15.2) -- *The subject of Assignment 1.*

#### XVII. Inspection Walkthroughs (Section 15.3).

- A. These are typically performed by a team composed developers, usability experts, and possibly users or user representatives.
- B. Per Nielsen, "*Cognitive walkthroughs involve simulating a user's problem-solving process at each step in the human-computer dialog, checking to see if the user's goals and memory for actions can be assumed to lead to the next correct action.*"
- C. Steps of a cognitive walkthrough:
  - 1. **Identify the characteristics of typical users.**
    - a. Describe product or prototype to users.
    - b. Define a specific task for users to complete.

2. *Convene designers and usability experts.*
3. *Walk through each action of the task, asking*
  - a. Will the correct action be evident?
  - b. Will the user notice of the correct action is available?
  - c. Will the user properly understand the response to their action, correct or incorrect?
4. *Record important information:*
  - a. Ideas about what causes user problems, and why.
  - b. Notes about design changes that may be needed, and other potentially relevant side issues.
  - c. A summary of the results.
5. *Revise the design to fix the problems.*

- D. The book makes the important observation that such walkthroughs should be *egoless*.
  1. I.e., designers should not try to defend their designs when it's clear there are problems.
  2. Usability experts should lose the "we're inherently more knowledgeable" attitude.
- E. The book also discusses an alternate form of *pluralistic walkthroughs*, which differ from the preceding cognitive walkthroughs in the following respects:
  1. Usage scenarios are developed as part of the process, not in advance.
  2. The analysis is involves more collaborative discussion than directed user observation.

#### XVIII. Predictive models (Section 15.4).

- A. Inspections are conducted without users, and without experts role-playing users.
- B. Rather, the inspection uses a formulaic model that predicts user behavior.

#### XIX. GOMS models (Section 15.4.1)

- A. It's an acronym for
  1. *Goals* -- what the user wants to achieve
  2. *Operators* -- user cognition and actions needed to achieve goals
  3. *Methods* -- learned procedures for accomplishing goals
  4. *Selection rules* -- used to choose which of multiple methods to select
- B. GOMS is a generic model, that does not predict specific user performance.

#### XX. The keystroke-level model (Section 1.5.2)

- A. In contrast to generic GOMS, this model provides actual numeric predictions of user performance.
- B. The model is based on the analysis of many empirical studies of actual user performance.
- C. The table on Page 709 lists some times for core tasks, including
  1. pressing a key
  2. pointing with a mouse
  3. clicking the mouse
  4. homing hands on the keyboard
  5. drawing a line with a mouse
  6. making a decision to do something (can vary significantly)
  7. system response time (can vary significantly)

- D. These numbers are used to add up the amount of time a particular task takes using a particular interface.
- E. The book provides a couple examples.

**XXI. Benefits (and limitations) of using GOMS (Section 15.4.3).**

- A. GOMS-style analyses do provides hard, quantitative data.
- B. Such analyses can in some cases lead to significant interaction design improvements.
- C. However,
  - 1. GOMS analyses are limited to a relatively small set of routine tasks.
  - 2. Do not model user errors.
  - 3. Nor do they model other cognitive, social, or environmental factors, such as
    - a. fatigue
    - b. distractions
    - c. multi-tasking
    - d. learning effects

**XXII. Fitts' law (Section 15.4.4).**

- A. Published in 1954 by Paul Fitts, it predicts the time it takes to reach a target using a pointing device.
- B. Can be used to determine where to place interface widgets on a display, and how big they should be.
- C. As the book's nutshell description says -- the bigger the target, the easier and quicker it is to reach.
- D. Specific HCI results based on Fitts:
  - 1. Don't have lots of tiny buttons crammed together.
  - 2. Things in the four corners of the screen are easy to reach (so why don't more UIs take advantage of this?)
- E. Some good interaction design results have been obtained by applying Fitts.
- F. And it's still with us, as evidenced by a full session in the 2008 SIGCHI entitled "*Fitt's Law Lives*".