

CSC 484 Lecture Notes Week 3, Part 1 (Re)Introducing Evaluation

I. Relevant reading.

- A. Textbook Chapter 12
- B. Papers of the fortnight:
 - 1. *"Storyboarding: An Empirical Determination of Best Practices and Effective Guidelines"* by Truong, Hayes, and Abowd, from the University of Toronto and Georgia Tech; Proceedings of the 6th conference on Designing Interactive systems, 2006, ACM.
 - 2. *"Developing Use Cases and Scenarios in the Requirements Process"*, by Maiden and Robertson, from the City University of London and the Atlantic Systems Guild; Proceedings of the 27th international conference on Software engineering (ICSE), 2005, ACM.

II. Class Schedule.

- A. First two weeks:
 - 1. Provide high-level introduction to ID, the process, and evaluation.
 - 2. These are covered in Chapters 1, 9, and 12 respectively.
- B. Week 3:
 - 1. Cover details of the requirements and prototyping process.
 - 2. Much of this is review material from CSC 308.
 - 3. This material is in Chapters 10 and 11.
- C. Weeks 4 and 5:
 - 1. Cover psychological, sociological, and cognitive aspects of ID.
 - 2. This is material not typically covered in depth in software engineering courses, at least here at Poly.
 - 3. It's covered in Chapters 2 through 5.
- D. Weeks 6 and 7:
 - 1. Summarize general paradigms of ID, i.e., ways of doing business.
 - 2. Cover statistical aspects of ID, specifically data gathering from users and analysis of the data.
 - 3. This is from Chapters 6 through 8.
- E. Weeks 8 and 9:
 - 1. Cover details of user evaluation.
 - 2. From Chapters 13 through 15.
- F. Week 10:
 - 1. Consider what's on the horizon and beyond.
 - 2. This will touch on topics from the research papers, particularly those from later in the quarter.

III. Assignment 2.

- A. The assignment topic is high-level storyboarding, what the book calls a "low-fidelity prototype" (Chapter 11, Section 11.2.3, Page 531).
- B. Hopefully, it will be the beginning of your term project.
- C. The culmination of the assignment is a two-day poster session in lab, during week 5.

D. See the writeup for details.

IV. Introduction to the class project.

- A. You will work in teams of approximately six members each.
- B. The default project theme is *productivity software for post-secondary education.s*
 - 1. "Default" means that you may work on a project in an entirely different area.
 - a. However, if your team has nothing else in mind, then post-secondary education is a reasonable application domain.
 - b. This is our collective, shared area of expertise, as students and faculty in this class.
 - 2. "Productivity" means it improves the educational lives of students and/or faculty.
- C. We will spend time in lecture on Wednesday discussing your (possibly tentative) project selection.
- D. Even if you are not sure of a project topic, you need a work area for the storyboarding task of Assignment 2.
- E. See the Assignment 2 writeup for further details.

Now on to the topic of user evaluation, as presented in Chapter 12.

V. Introduction to evaluation (Section 12.1).

- A. The purpose of any form of evaluation is to collect information about users.
- B. There are multiple possible methods to do so.
- C. In Assignment 1, you approached evaluation as a team of expert users, evaluating fully finished products.
 - 1. We did this as an immediately accessible form of evaluation, as a "warm-up" exercise for the class.
 - 2. Later in the quarter, you will perform a laboratory-based evaluation of actual end users, who may or may not be experts in the domain of product use.

VI. The "why", "what", "where", and "when" of evaluation (Section 12.2).

- A. *Why?*
 - 1. Check that users can do something useful with a product.
 - 2. Check that they like it.
- B. *What?*
 - 1. Evaluate the product itself.
 - 2. Evaluate domain-specific attributes, including performance, aesthetics, physical characteristics.
- C. *Where?*
 - 1. Evaluate in a controlled laboratory setting.
 - 2. Evaluate in natural settings of use.
- D. *When?*
 - 1. Evaluate at any appropriate stage of the development process.
 - 2. Do concept evaluations at the beginning of developing a brand new product.
 - 3. Evaluate specific new features when a product is being upgraded.
 - 4. Evaluate a finished product, including for standards compliance.

VII. Evaluation terminology (alphabetically listed in Box 12.1).

- A. *Analytic evaluation* -- An approach using heuristics, walk-throughs, or models, without actual end users.
- B. *Controlled experiment* -- Evaluation of actual users, in a controlled laboratory setting.
- C. *Field study* -- Evaluation of actual end users, in their natural environment.
- D. *Formative evaluation* -- Done during design, to ensure continued user satisfaction.
- E. *Heuristic evaluation* -- Done using well-known guidelines, embodying knowledge of typical users.
- F. *Predictive evaluation* -- Done with theoretical models, to (attempt to) predict user performance.
- G. *Summative evaluation* -- Done when a design is complete, in particular to assess standards compliance.
- H. *Usability lab* -- A facility designed specifically for usability studies.
- I. *User study* -- Any kind of user study, at any stage development.
- J. *Usability testing* -- A quantitative evaluation study.
- K. *User testing* -- Evaluation of users performing specific tasks.

VIII. Approaches and methods (Section 12.3).

- A. There are three widely-used approaches -- usability testing, field studies, and analytic evaluation.
- B. They can be used at various stages of product development, separately or in combination.

IX. Approaches (Section 12.3.1).**A. Usability testing**

- 1. Done in a lab or similar setting.
- 2. The environment is well controlled by the evaluators.
- 3. Test subjects must focus on the tasks at hand and not be interrupted, e.g., by phone calls or other typical day-to-day activities.
- 4. Quantifying user performance is an important aspect of usability testing.
 - a. All users are given the same tasks, and measured in specific ways.
 - b. When such tests are conducted over a single product's full life span, this is a form of regression testing, in the software engineering sense.
 - c. I.e., the same tests are used with successive product releases, to ensure that a core set of tasks can be performed in a new release at least as effectively as in a preceding release.
 - d. This has been called "usability engineering".

B. Field studies

- 1. In contrast to usability testing, field studies are done in users' natural settings.
- 2. Subjects are observed in an unobtrusive manner, recording their activities in different forms, including audio and video if possible.
- 3. Subjects may be asked to fill out questionnaires about their experiences.

C. Analytic evaluation

- 1. This can be done using *heuristic-based walkthroughs* or *models*.
- 2. This form of testing does not involve actual end users, where "actual" means the intended user population or its designated representatives.
- 3. Rather, it is conducted by product developers, most typically domain experts.
- 4. Heuristics are developed to characterize typical user behavior.
 - a. They can be based on common-sense knowledge in a particular domain.
 - b. They also involve other general or specific guidelines of product usage.

5. Models are scientific attempts to characterize certain types of measurable user behavior.
6. E.g., Fitt's law predicts the time it takes to reach a target using a pointing device.
7. *Cognitive walkthroughs* are a form of modeling that involve simulating a users problem-solving process, to determine how users will interact with a product.
8. Overall, analytic evaluation can be useful, but is never a replacement for actual end user testing.

X. Methods (Section 12.3.2)

A. The main methods employed in evaluation are:

1. *Observing users.*
 - a. In a lab.
 - b. In the field.
 - c. With direct evaluator contact and/or indirect recording.
 - d. Recording in various ways, including audio, video, and product instrumentation.
2. *Asking users their opinions*
 - a. Individual in-person interviews, with note taking.
 - b. In group meetings and discussion sessions.
 - c. Using questionnaires.
3. *Asking experts their opinions.*
4. *Testing users' performance.*
5. *Modeling users' performance.*

B. Table 12.1 on Page 594 is a useful summary of the different evaluation approaches and methods.

XI. Case studies (Section 12.4).

A. The book provides six short case studies to illustrate the use of various evaluation methods.

B. Here is an overview of the evaluation methods we will be employing in 484:

1. Heuristic evaluation of an existing project in Assignment 1.
2. Informal interviews and questionnaires for project ideas in Assignment 2.
3. Lab-based usability study of an existing or new product in Assignment 3.