Tutorial Proposal

Title

Advanced Usability Testing Methodology.

Duration

Full day.

Learning Objectives

Most usability professionals learned their skills by reading books on usability testing or by watching someone else prepare and conduct tests. But how do we know that the particular way we learned to do usability testing is still the best way? Until the Comparative Usability Evaluation (CUE) project, there was little opportunity for usability professionals to objectively compare their various approaches. As it turns out, there are some striking differences in effectiveness.

This tutorial gives rare insights into the practical doings of usability professionals, normally shrouded behind walls of confidentiality. This tutorial is not based on the personal opinions of one or two instructors but on the practical accumulated experience - both good and bad - of the eleven professional usability labs who participated in the CUE-1 and CUE-2 projects, conducting controlled usability tests in realistic, industrial settings.

In this tutorial participants will learn:

- What constitutes quality in usability testing.
- Options available in planning a usability test, and how the choices may affect the outcome.
- What makes a good usability test scenario, and how test scenarios impact the quality of a usability study.
- What a usability problem is, and how usability problems are identified.
- How many test participants are required to find for example 85% of the usability problems in a product.
- Required characteristics of a usability report to assure its communicative value.
- Novel communication techniques that are vastly superior to traditional usability problem communication through paper reports and video tapes.

Insight from the professional lab studies will enable participants to assess and improve their abilities in usability test planning, scenario design, usability reporting and usability problem communication.

Background

CUE-1

In early 1998 four professional usability labs performed independent tests of a Windows calendar management application, Task Timer for Windows. The results of this comparative study was published in [Molich et al. 1998]. The study is called "Comparative Usability Evaluation 1", CUE-1.

The participating usability labs were:
- HFRG (Human Factors Research Group), University College Cork (Ireland)
  Jurek Kirakowski
- National Physical Laboratory (UK)
  Nigel Bevan and Ian Curson
- Rockwell Software (USA)
  Scott Butler
- Sun Microsystems, JavaSoft Division (USA)
  Erika Kindlund and Dana Miller.

All usability tests were carried out by experienced usability professionals employed by the labs.

The study showed that there were remarkable differences in approach, reporting, and findings between the labs. The most interesting result was perhaps that while a total of 141 usability problems were uncovered by the four labs, only one problem was reported by all four labs, and only one other problem was reported by three labs. Eleven problems were reported by two labs. Each of the remaining 128 problems were reported by only one of the labs.

We also found that the usability reports generated by the labs differed considerably from each other. They also differed from the recommendations presented in some of the recognized textbooks in the field, like [Dumas and Redish 1994] and [Rubin 1993].

The results of CUE-1 were presented at a panel discussion and in a refereed paper at the annual Usability Professionals' Association conference in 1998 and are reported in [Molich et al. 1998].

CUE-2

The CUE-1 study generated considerable interest. Therefore, seven other professional usability labs decided to undertake another, similar study in the fall of 1998. This study is called "Comparative Usability Evaluation 2", CUE-2. Some of the participants from CUE-1 agreed to serve as advisors for this new round of tests.

The setup for the study was similar to that used in CUE-1. The main improvements were:
- Access to developers during the usability test through monitored exchange of questions and answers by e-mail
- Test of a state-of-the-art web-site (instead of a four year old Windows application)
- Participation by anonymous student teams in order to investigate whether results from usability testing done by professional labs is clearly superior to that done by good student teams

The participating usability labs were:
- Kommunedata (Denmark)
  Klaus Kaasgaard
- Networkers (Denmark)
  Lars Schmidt
- NovaNET Learning, Inc. (USA)
  Joseph Seeley
- P5 (The Netherlands)
  Wilma van Oel
- SGI (USA)
  Barbara Karyukina
- Sun Microsystems (USA)
  Meghan Ede
- University of Maryland, Dept. of Psychology (USA)
  Kent Norman

The application used in the CUE-2 usability test was the web-site www.hotmail.com. Hotmail provides free e-mail service to users of the Internet.

Results of CUE-2 were presented in a panel at CHI '99. This presentation and other CUE-papers can be found on www.dialogdesign.dk/cue2.htm

The study confirmed the results from CUE-1 that there were remarkable differences in approach, reporting, and findings between the labs.

Rolf Molich conceived the original idea of the comparative usability evaluation and coordinated both CUE-1 and CUE-2.

Industry Response

The results of both CUE-1 and CUE-2 were presented to industry professionals at two international conferences, UPA'98 and CHI'99, and generated further discussion towards issues surrounding usability engineering. Industry experts offered opinions and interpretations of the CUE data. The experimental data, in conjunction with industry feedback to the CUE exercises, offer valuable insight into such questions as:
- Is usability engineering a science. If so, what makes it one? If not, is it still valuable to the design process?
- Can usability studies generate reproducible results? Under what conditions?
- How do individual usability professionals differ in their approach and how do they justify these differences?
- Are certain approaches more successful than others, and how is this success measured?

Material Covered by the Tutorial

The main material covered by the tutorial is:
- Overview of the Comparative Usability Evaluations (CUE) and related work.
- What is quality in usability testing?
- Options available in planning a usability test, and how planning choices may affect the outcome of the test.
- How usability test scenarios impact the outcome of a usability study.
- What makes a good usability test scenario?
- What is a usability problem? How were usability problems identified in CUE?
- Required characteristics of a usability report to assure its communicative value.
- How the usability report can determine the impact your findings have on directing change in the product development cycle.
- Reactions from customers (Time/system and Hotmail) to the CUE tests.
- Reactions from participating CUE test teams to the comparative evaluation of their work.
- General lessons learned from CUE-1 and CUE-2. See the section "Industry Response" above.

The material covered by the tutorial is described in more detail in the section Schedule below.

Justification for a CHI Audience

This tutorial is highly relevant for a CHI audience because it gives a rare insight in the practical doings of usability professionals. This tutorial is not based on the personal opinions of a lecturer but on the practical accumulated experience – both good and bad – of eleven professional usability labs and two student teams.

The tutorial is particularly relevant for practitioners with moderate or large experience in running usability tests under industrial conditions. Through non-offensive exercises, this tutorial will show them their strengths and weaknesses in one of the core processes of the usability profession.

The tutorial is also relevant for academics who can use the considerable practical experience that is communicated by this tutorial to improve their teachings. Teachers may also be able to use the applications and the usability test scenarios for student exercises in their own courses.

Although this tutorial is not primarily aimed at novice practitioners, novices can get a good look at a number of realistic problems and suggested solutions in practical usability testing.

How the Tutorial will be Conducted

The tutorial will consist of lectures, discussions and practical exercises as described in detail in the following section.

Schedule

This proposal is for a full day tutorial. The tutorial is divided into four modules of 90 minutes each.

Module 1   **Introduction**
- Overview of tutorial
- What is quality in usability testing?
- The Comparative Usability Evaluations (CUE-1 and CUE-2)
- Survey of related work
- The usability lifecycle (taxonomy)

Exercise 1: Usability Test Requirements

Participants will review the requirements specification for the usability tests in CUE-2, and create a checklist for important items for such a requirements specification.

Designing a Usability Study
- Methods used in CUE: Inspection, Inquiry and Testing
- The Methodological Effect: How results varied based on methodology used in CUE
- Discount test methods

Module 2   **Exercise 2: Review Usability Test Tasks**

Participants will review some of the usability test scenarios used in CUE-2 for testing www.Hotmail.com. Participants will be encouraged to find problems in the experts' test scenarios.

Tasks and Scenarios
- The Scenario Effect: How usability test scenarios impact the outcome of a usability study
- Common problems in tasks and scenarios
- What makes a good usability test scenario?

Module 3   Identifying and Describing Usability Problems
- What is a usability problem?

Exercise 3: Usability Problem Description
Participants will review problem descriptions in the CUE reports and identify desirable and questionable elements of a usability problem description. Based on a specific, non-trivial example, we will also discuss guidelines for suggesting usability problem corrections.

Communicating Results
- What is effective usability problem communication?
- Various means of communicating results (other than reports)
- The Report Effect: How the usability report can determine the impact that findings have on directing change in the product development cycle
- Checklists: Required characteristics of a usability report
- An example of a usable usability report (for www.TowerRecords.com)
- The KJ Method: Quick and effective usability problem communication without reports

Communication-in-the-Large: The Politics of Usability

Module 4   Tips and Tricks
Participants will select two of the following topics for an in-depth discussion with other experienced professionals:
- Finding good test participants
- Performing tests
- Testing intermediate and advanced users
- International usability testing
- Assessing the quality of a usability consultancy

Conclusion
- Critical comments on CUE
- Reactions from participating CUE test teams to the comparative evaluation of their work ("The psychology of usability testing")
- General lessons learned from CUE
- Why exhaustive usability testing of non-trivial products is impossible
- Prevention is better than cure
- Quality through humility in usability testing

Materials

The instructor will use a computer screen projector to present the lecture material.

The tutorial notes will include
- Abstract.
- Agenda.
- Copies of all the overhead slides used by the instructor with space on each page for attendees to take notes.
- Checklists for conducting usability tests, setting up a usability report and reporting individual problems with particular focus on findings that are not covered by the standard textbooks [Dumas and Redish 1993] and [Rubin 1994]
- Excerpts from selected, anonymized reports from the Comparative Usability Evaluation studies. These reports provide an excellent overview of the state-of-the-art within usability testing and reporting.
- Copies of the refereed papers that summarize the results of the Comparative Usability Evaluation studies on which the tutorial is built.
- Instructor biography.
- Annotated bibliography.

The anonymous CUE-reports on which this tutorial is based can be found on www.dialogdesign.dk/cue2.htm

Appendix A contains sample handouts from module 3.

Instructor

Rolf Molich, DialogDesign

Rolf Molich owns and manages DialogDesign (www.dialogdesign.dk), a small Danish usability consultancy. Rolf conceived and coordinated the comparative usability evaluation studies CUE-1 and CUE-2 where 11 professional usability teams and two student teams tested the same application. Rolf was a principal investigator in the Nielsen Norman Group's recent large scale usability test of 20 US E-Commerce websites involving more than 60 users. Rolf has worked with usability since 1984; he is the co-inventor of the heuristic inspection method (with Jakob Nielsen) and the author of the best-selling Danish book *User Friendly Computer Systems* of which 25,000 copies have been sold.

Rolf is an experienced speaker. His 24 tutorials at the recent NN/g World Tour attracted about 1,000 participants. The overall average participant evaluation of his tutorials was 4.46 on a five point scale.

History

This tutorial has previously been presented at CHI 2000 and at the Nielsen Norman Group World Tour.

CHI 2000 tutorial

An early version of this tutorial was presented at CHI2000 as tutorial #21 "Improving Your Skills in Usability Testing" on 3. April 2000 in Den Haag, The Netherlands. This tutorial was conducted by Rolf Molich and Erika Kindlund

The following is a summary of the evaluation forms returned by the attendees of the CHI2000 tutorial.

The tutorial had 67 responding attendees

The following items 6 through 15 and 17 from the standard CHI2000 evaluation form are reported as averages on a 7 point likert scale where 1=strongly disagree and 7=strongly agree

6. The description of this tutorial in the Advance Program was accurate   5.1
7. The tutorial notes supported the presentation of the tutorial   6.0
8. The tutorial notes will be a valuable reference   5.5
9. The slides/overhead transparencies were of high quality   5.8
10. The instructor(s) made effective use of audiovisual media besides
    slides and overhead transparencies   5.6
11. The tutorial offered valuable interaction with the instructor(s)   5.6
12. It would have been difficult to obtain the information in this
    tutorial from somewhere else   5.0
13. The material was well-organized.   5.6
14. The tutorial was worth the time and money.   5.6
15. The tutorial should be offered again next year.   5.7

16. The technical level of this tutorial was:
    - too introductory according to 9 attendees
    - about right according to 57 attendees
    - too advanced according to 1 attendee

17. The material that Erika Kindlund presented was relevant and important 5.8
    Erika Kindlund communicated the material effectively.   5.2

    The material that Rolf Molich presented was relevant and important.   6.0
    Rolf Molich communicated the material effectively.   6.0

Nielsen Norman Group Tutorials

After CHI2000 Rolf Molich has presented this tutorial 12 times as part of the Nielsen Norman Group World Tour from November 2000 to April 2001.

On a five point scale, the average overall satisfaction with this tutorial was 4.46. The overall satisfaction slowly increased during the World Tour, probably reflecting the constant improvement of the tutorial. The overall satisfaction with the final tutorial in April 2001 on which the proposed CHI2002 tutorial is based was 4.57.

The tutorial has been continually revised in response to constructive criticism from participants. The most significant changes as compared to the CHI2000 tutorial are:
- More time to discuss personal experiences and exchange tips and tricks (module 4).
- More emphasis on the political aspects of usability testing.
- More emphasis on novel problem communication techniques like the KJ-method.
- Requirements specification exercise considerably improved based on participant feedback.
- Scenario examples considerably improved based on participant feedback.
- Live usability test omitted.

References

[Molich et al. 1998]   Rolf Molich, Nigel Bevan, Scott Butler, Ian Curson, Erika Kindlund, Jurek Kirakowski, and Dana Miller: *Comparative Evaluation of Usability Tests*; Proc. Usability Professionals Association 1998 (UPA98), pp 189-200.

[Molich et al. 1999]   Rolf Molich, Klaus Kaasgaard, Barbara Karyukina, Lars Schmidt, Meghan Ede, Wilma van Oel, Meeta Arcuri: *Comparative Evaluation of Usability Test;* CHI99 Extended Abstracts (ACM Press 1999), pp 83-84.

[Dumas and Redish 1993] Joseph S. Dumas and Janice C. Redish: *A Practical Guide to Usability Testing*; Ablex 1993

[Rubin 1994]   Jeffrey Rubin: *Handbook of Usability Testing*, John Wiley 1994

Appendix A. Sample Handouts from Module A

Advanced Usability Testing Methodology

Communicating Results

Rolf Molich
DIALOGDESIGN

Primary Purpose of a Usability Test

● The primary purpose of a usability test is to cause beneficial improvements to the user interface.

● The primary purpose of a usability test is NOT to write a good usability report.

Several CUE-teams and independent observers have made these important points.

Mini Exercise E-1
Communicating Test Results

Brainstorm:

- Suggest ways (other than usability reports, video tapes and video highlight tapes) to communicate test results.

?

Mini Exercise E-1
Communicating Test Results

Our suggestions:

Get buy-in from developers.

- Work with your developers to define test tasks and user profiles.

- Ask developers to watch tests.

- Communicate results by email ("Real programmers don't read reports").

- Build consensus, for instance through the use of the KJ-Method.

Consensus Building (1/2)

The KJ-method (by Kawakita Jiro):

- Ask developers to watch usability test and take notes.

- Immediately after the last usability test get the developers together.

Consensus Building (2/2)

- Ask each participant to write down each major observed problem on a colored index card.
- Put all index cards on a large, sticky board. No discussions!
- Read each others cards silently. Add additional problems.
- Sort problems by area. Eliminate duplicates if total agreement.
- Name each group.
- Vote for most important problems. Each participant has ten votes (x'es).

Consensus Building - Reference

KJ Method, reference:

● Tara Scanlon
  Getting to Consensus Quickly
  Eye for Design, January 1999, pp. 4-5.

Eye for Design is published bi-monthly by
User Interface Engineering, www.uie.com

Eye for Design has lots of other fine
articles about practical usability testing.

Recommended!

Communication - The Politics of Usability

"God is on our side"

- but you can help him.

You must sell, remove obstacles, motivate,
implement and strengthen usability.

Communication - The Politics of Usability

- Make the consequences of ignoring usability visible.
- Avoid opinions - test!

Communication - The Politics of Usability

- Don't fall into the tar pit of opinions. Be careful with focus groups and heuristic inspections.
- Developers and designers are skilled in discussing opinions. Opinion wars only have losers.
- There is no magic answer to the question "Why are your opinions better than mine?"
- You become a new and interesting player if you continue to sell your ignorance by insisting that only user testing has the right answers.

Communication - The Politics of Usability

- Build alliances.

- Avoid confrontations.

- Pick your battles; you can't win them all
  (and even if you could, it would be unwise).

- Position yourself as an allied,
  not as an enemy or police function.

- Build trust by being completely open. Demystify.

- Document and sell successes to management.

Recommended Readings

The politics of usability testing are described in:
- Deborah Mayhew
  The Usability Engineering Lifecycle (chapter 18),
  Morgan Kaufman Publishers, 1999.

Also appeared in:

- Deborah Mayhew
  Strategic Development of the Usability Engineering Function,
  interactions, volume VI.5, September 1999.

Usability Test Reports

Team	A	B	C	D	E	F	G	H	J
# Pages	16	36	10	5	36	19	18	11	22
Exec summary	Y	Y	N	N	N	Y	N	Y	Y
# Screen shots	10	0	8	0	1	2	1	2	0
Severity scale	2	2	3	1	2	1	1	3	4

Usable Usability Report

- Keep it short - 12-15 pages.
- One page executive summary.
- List comments with type, severity, #users.
- Avoid opinions.

Usability Results

Team	A	B	C	D	E	F	G	H	J
# Problems	32	149	18	10	67	76	41	18	25
# Highest severity	9	26	6	-	7	-	-	0	0
# Exclusive	13	105	4	0	33	36	19	10	12
# Positive findings	0	8	4	7	24	25	22	4	6

Exclusive: Problems reported only by this team

Include Positive Findings in Your Reports

Why?

Include Positive Findings in Your Reports

Positive findings are important

- To make the report trustworthy.

- To make developers more open to criticism.

- To ensure that developers do not remove features that users appreciated.

Include Positive Findings in Your Reports

One report started by saying:

"Generally, the users were very happy about Hotmail."

The rest of the report contained more than 30 problem descriptions without any positive findings to substantiate the initial claim.

Observed Problems (1/4)

- Report too long. Too many problems reported.

- No executive summary.

- No screenshots.
  - Dull report.
  - Limited value for others.

- Few user quotes.
  - Good quotes "make you feel the user's pain"

- Harsh words (Even developers have feelings!).
  Example: "Default window size does not display
  Add/Edit/Delete buttons!!!!"

Observed Problems (2/4)

- Some reports: No clear distinction between
  - Personal opinions,
  - Expert opinions,
  - User opinions,
  - User findings.

- Insufficient information to reproduce test.
  Put test scenarios, participant profiles, questionnaires,
  interview questions, etc. in an appendix.

Observed Problems (3/4)

- Unattractive, unprofessional layout.

- Incomprehensible problem descriptions.

- Descriptions that were a conglomerate of 2-6 more basic problems.
  A "basic" usability problem is a problem that can be corrected without affecting other problems.

Observed Problems (4/4)

- Vague descriptions that would probably not help a design team.
  Example: "Severe problem: Terminology was often confusing, especially when different terms referred to similar features or the same feature could be accessed by different terms."
  Solution: Provide examples.

Advanced Usability Testing Methodology

INSTRUCTOR
Rolf Molich – DialogDesign

BENEFITS
Compare your own approach to usability testing with those used by eleven professional labs during controlled usability tests in realistic, industrial settings. This tutorial gives a rare insight in the practical doings of usability professionals.

ORIGINS
Presented at CHI 2000 and at the Nielsen Norman Group World Tour.

FEATURES
- Gather insight from practical examples of usability work done by other professional labs – both good and bad
- Improve your abilities in usability test planning, scenario design and usability reporting
- Improve your abilities in identifying usability problems
- Learn about novel usability problem communication techniques that are vastly superior to traditional techniques (paper reports and video tapes).

AUDIENCE
Usability professionals who have conducted a few tests and would like to improve their skills in test planning, scenario design, problem identification, and problem communication. Those with more experience may be familiar with some material but will still benefit by learning from other usability professionals. Although this seminar is not intended as an introduction to usability testing, past participants with no testing experience have rated it highly.

PRESENTATION
Lecture segments interspersed with exercises and discussions.

INSTRUCTOR BACKGROUND
Rolf Molich owns and manages DialogDesign, a small Danish usability consultancy. Rolf coordinated the comparative usability evaluation studies CUE-1 and CUE-2 where eleven professional usability labs tested the same application. Rolf has worked with usability since 1984; he is the co-inventor of the heuristic inspection method (with Jakob Nielsen).

Requirements List for Tutorial

Title

Advanced Usability Testing Methodology

Instructor

Rolf Molich, DialogDesign

Requirements List

For instructor:

- XGA projection of instructor's portable computer display (768 x 1024 pixels).

- Table

- Wireless microphone

- Flip-chart with pens of various colors.

For tutorial participants:

- Pens and pencils.

- One paper notepad per tutorial participant.