

Usability lab at Oracle

<http://flickr.com/photos/52056151@N00/1846245768/>

I214: Usability Test Design

February 14, 2013

What is usability?

Learnability

How easy is it for users to accomplish basic tasks the first time they encounter the design?

Efficiency

Once users have learned the design, how quickly can they perform tasks?

Satisfaction

How pleasant is it to use the design?

Memorability

When users return after a period of not using it, how easily can they reestablish proficiency?

Errors

How many errors do users make, how severe are these errors, and how easily can they recover from the errors?

Usability how-to: What we will cover

Plan

- Define goals!
- Decide on scope of study
- Design method
- Data collection
- Measures
- Users

Meeting 1 and 2

- Study Design
- Questionnaire construction

Set up study

- Recruit users
- Develop materials

- Recruiting
- Observation
- Interviewing

} Earlier meetings

Conduct study

Record

- Analyze findings
- Report findings

Meeting 3

- Analysis
- Reporting



Types of usability testing

Exploratory

Testing preliminary design concepts

Assessment

How well does this design work?

Comparison

One design against another

Validation

Certify usability late in development

Does it meet standards, benchmarks

The classical usability test

Participants are or represent real users

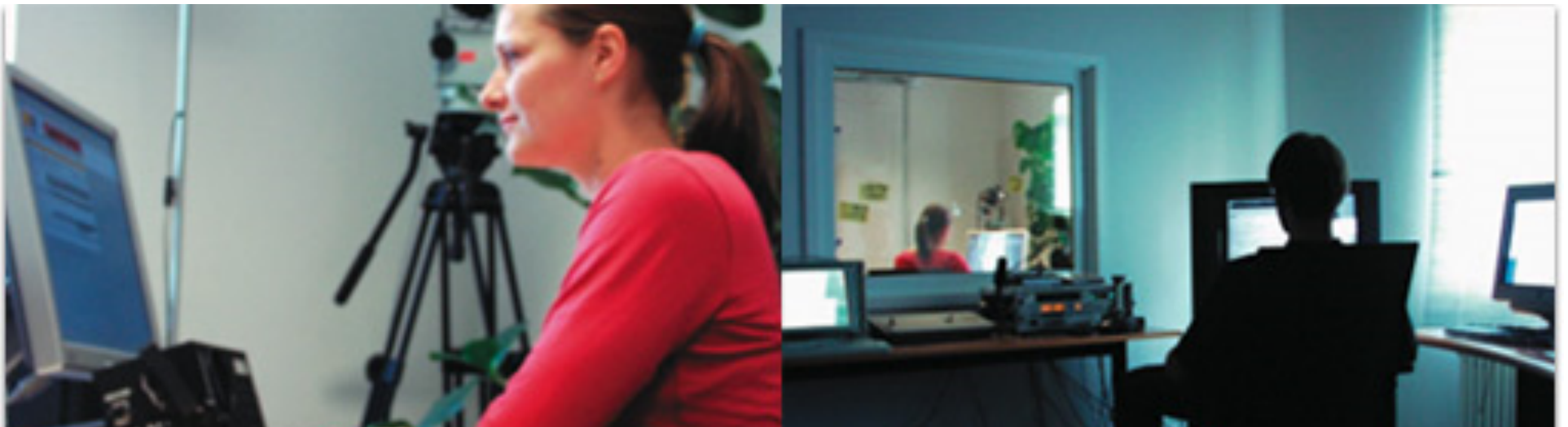
Tasks are typical, setting is controlled

Measures performance quantitatively

Observations are recorded with little interference

(Optional) **Debrief**

Via questionnaire, interview, and/or focus group,





Usability mixes methods

Quantitative

Measure system
and user
performance

errors, # keystrokes
needed, time required,
recall/precision

Qualitative

Interview-based
examination of individuals'
actions, problems,
confusions, concerns

Classical usability testing is rooted in experimental design

Principles

Validity

Internal to the test

External, to real world

Reliability



Methods

Designed to control for influence of factors *other than what's being tested*

The goal of classical usability testing: Maximize the likelihood of a valid and reliable test. **But how?**



MINIMIZING THREATS TO VALIDITY AND RELIABILITY

Controlling participant conditions

History

Avoid recruiting participants with conflictual experiences or backgrounds

Selection

Number

Representative of user population

Randomly assigned to treatment groups (if multiple)

NOT professional testers

Avoid selecting participants who are likely to be
anxious or fearful

Controlling study design

Tasks

Controlled/limited (internal validity)

Typical (external validity)

Minimize feeling of being a “test subject”

Maturation of experience with systems and tasks

Caution about re-using subjects

Varying order of activities, tasks, of system designs tested (if >1)

Be wary of novelty effects

Beware of fatigue, discouragement, boredom...



Controlling test conditions

Test environment

replicated for each participant
but sets people at ease

Experimenter

Neutral demeanor
Minimize interactions
Use a script
Consider experimenter
characteristics vs participant

Instrumentation

use identical interfaces,
browsers, survey forms
Purge caches, history, and
autocompletes

Mortality

Note and avoid dropouts

Managing sources of error

AN COMPARATIVE TEST EXAMPLE

Example: Cellphone test

The participants were selected from the 41 subjects responding to the open online survey.... Nineteen participants were selected for the comparison test, and assigned to one of two groups. We generated matched groups that had approximately equal numbers of each gender and of varying experience with a Nokia TM cellular phone.

Table 1. Characteristics of Test Participants

	Group 1 (n=9)	Group 2 (n=10)	Total (n=19)
Male	5	5	10
Female	4	5	9
Nokia TM user	5	6	11
Non-Nokia TM user	4	4	8

Huang et al, 2006
Empirical Evaluation of a Popular Cellular Phone's Menu System: Theory Meets Practice

Example: Cellphone test

Repeated measures (within-subjects) design

Group 1	cellphone A	cellphone B
Group 2	cellphone B	cellphone A

Example: Cellphone test

Compensate for cellphone maturation effect

Group	Tasks	
	Task set X	Task set Y
1	cellphone A	cellphone B
2	cellphone B	cellphone A

Example: Cellphone test

Compensate for cellphone maturation effect *and* growing familiarity with tasks

Group	Tasks	
	Tasks X	Tasks Y
1	cellphone A	cellphone B
2	cellphone B	cellphone A
	Tasks Y	Tasks X
3	cellphone A	cellphone B
4	cellphone B	cellphone A



Managing sources of error

SETTING TYPICAL AND CONTROLLED TASKS

Example: Cellphone test

“Five representative tasks were chosen **based on the online survey** of participants’ most frequent use of handset operations [...] In addition, tasks were to

(1) be **representative** of the whole population of tasks, and

(2) **capture some of the major modification** differences between Model A and Model B. ”

Tasks

1. Check received calls.
2. Find the wireless Internet access.
3. Find the option, “Welcome Note”.
4. Turn on vibrating alert.
5. Set the phone on the silent mode.



Managing sources of error

MEASUREMENT

Match measurement to project-specific goals and concerns

Performance

Actions and behaviors

How many, how long, what %

Time to task completion

Number of errors

Type of errors

Number of attempts

Eyetracking results

Subjective

Perceptions, opinions, judgments

Users AND Observers

Still requires rubrics

Contextualized: when, where, what were they doing



Sources of measures

Perceived problems and strengths

Users (interviews, forums, etc)

Internal stakeholders

Partners

Heuristics*

E.g. “easy to use”:

How many times did users refer to “help,” act confused?

How many screens did user have to look at to do X?

Ask users at end of test: “How easy...?”



Example: Cellphone test

Measures

1. The time to complete each task.
2. The number of attempts to complete each task.
3. Task success rate.
4. Number of and types of errors:
Observations and Comments: note when participants had difficulty, when an unusual behavior occurred, or when a cause of error became obvious.

Noncritical Error: a participant made a mistake but was able to recover during the task in the allotted attempts.

Critical Error: a participant made a mistake and was unable to recover and complete the task successfully.

Example results: Cellphone test

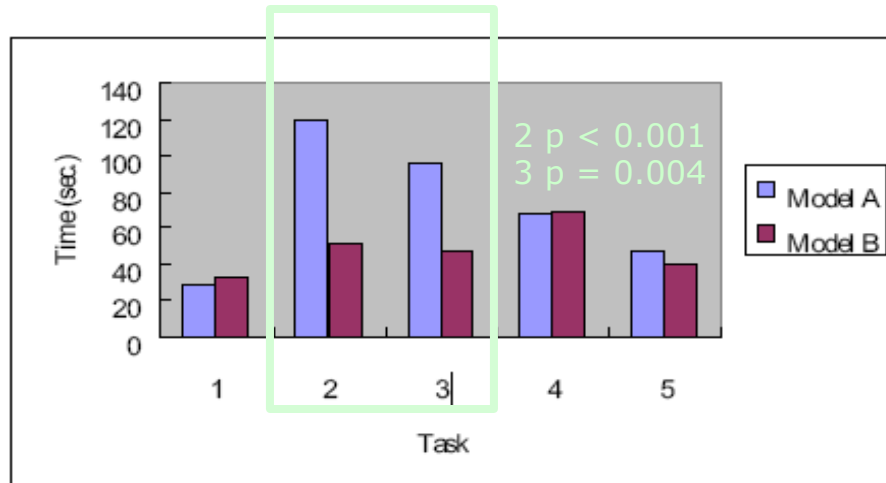


Figure 3. Average Time Span for Tasks in Model A and B

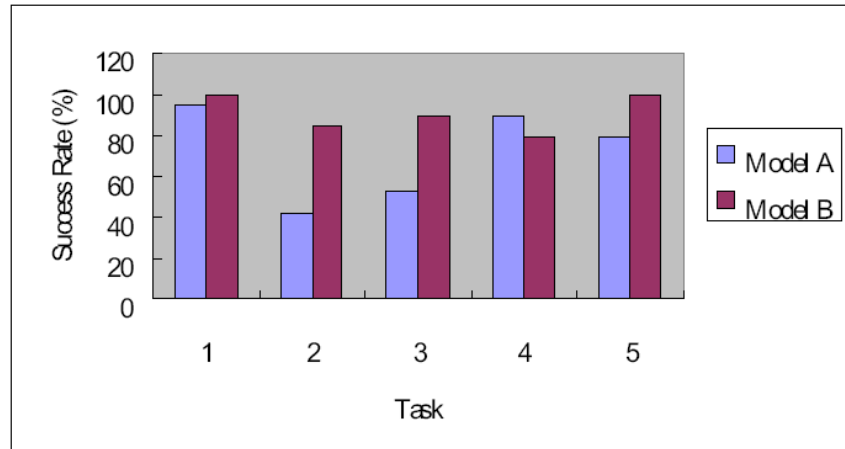


Figure 5. Success Rates for Tasks in Model A and B

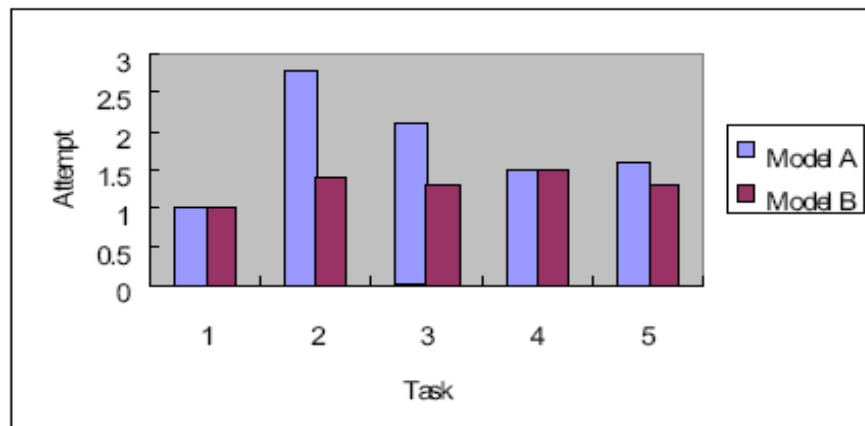


Figure 4. Average Attempts for Tasks in Model A and B

“Although people tended to take longer and make more attempts on tougher tasks, neither Task 2 nor Task 3 was the most complex task among the tasks.”

Benefits of formal usability testing

Benefits of formal usability testing

Clear results

Clear feedback

Added credibility

in quantitatively-oriented organizations, with quantitative professions
when client, designers can see and hear users

Limits to usability testing

Limits to usability testing

Unrepresentative conditions

Unrepresentative tasks?

- Limited kinds of tasks amenable to testing

- Short time period

Unrepresentative users?

- Limited number, range of users

- Often novice users (e.g., for a new interface)

- Testing effects: people do their 'best' when being observed

Limited observation opportunity

Typical methods combinations

Interviews, surveys (to find areas of concern) followed by testing (to explore areas of concern)

Testing followed by questionnaires, interviews, or focus groups to improve understanding of test findings

Testing based on results of internal inspections
once possible problems are uncovered, use testing to examine more rigorously