

Task Oriented User Evaluations

- Not an experiment! Don't always have a specific hypothesis I want to test
- Might well have a new system that I want to get reaction from real users

Or

- make sure that real users don't have any problems in using

Or

- Might want to see how it compares to my usability criteria

Task Oriented User Evaluations

Criteria

- Proportion of tasks successfully completed (effectiveness)
- Time to complete each task (efficiency)
- Errors made (type and number, efficiency, time/error trade off)
- Time required to obtain 95% error free performance (learnability)
- Proportion of tasks successfully completed after a certain time from learning to use the system (memorability)
- Perceptions of the system – measured on Likert scales (satisfaction)

Identifying Problems

- If the task is too complex to do while making a verbal protocol or there's some other reason why you don't want/can't use that technique
- You can use a retrospective verbal protocol instead
 - You record the participant doing the task and then play back the recording, asking them to talk through their thoughts/problems etc
 - Has the added advantage that if the commentary is complex, you can pause the recording and discuss things with the participants and then continue

Task Based User Evaluation Problem Summary

Problem Description	Severity	Solution
User can't understand if their pictures are secure	Catastrophe	Redesign!
Overwhelming number of items in privacy settings	Major	Categorization of links into appropriate groupings Aggregate some settings
Privacy settings hidden under account menu not in profile area	Major	Make link under profile more obvious to the user
Privacy settings under table labelled "notifications" - very strange labelling	Minor	Change label to Notifications and Privacy
Page "My profile" not the same as the menu item selected - disorienting	Cosmetic	Change menu label

SHNEIDERMAN'S 8 GOLDEN RULES

1. Strive for consistency
2. Cater to universal usability
3. Offer informative feedback
4. Design dialogs to yield closure
5. Prevent errors
6. Permit easy reversal of actions
7. Support internal locus of control
8. Reduce short-term memory load

SHORT SUMMARY

- Lab-based studies
- Need for statistics
- Research vs design

USABILITY ENGINEERING

- Build interfaces
- Measure performance
- Improve measures
- Everybody happy...?

MEASURES

- Efficiency
 - Task time
- Effectiveness
 - Tasks completed
- Satisfaction
 - QUIS, SUS

PROBLEM

- People vary
- How do you know your design is *the cause of differences?*
- Eg testing human memory

CAR PARK MONITORING SYSTEM

- What measures should we use to evaluate the system?

EXPERIMENTAL APPROACH

- Borrowed from Psychology
 - Laboratory(-based) studies
 - Quantitative methods
- Usability testing
 - Lab-based evaluation

USABILITY TEST DESIGN

- Independent variable
 - control vs new design 1 vs new design 2
- Dependent variable
 - Task time
- Eliminate confounds
- See if designs are different

USABILITY TEST ARGUMENT

- Hypothesis: design I is *quickest* to use
- Time people using each design
- Average task time on each design
- Use statistics to see if differences are real
- Real differences are due to design

DEVISE A USABILITY TEST OF THE CAR MONITORING SYSTEM

- Evaluate the effectiveness of the visualisation of the rankings
- Hypothesis?
- Independent variable?
- Dependent variable?

ELIMINATING CONFOUNDS

- Random allocation
- Multiple tasks
- Ordering effects

STATISTICS (IN ONE SLIDE!)

- Variation vs differences
- Impossible to be sure
- Need predictions
- Probability machine
- Unlikely predictions are significant

BENEFITS OF TESTS

- Lots of experience out there
- Focused questions
- Rigorous
- Quantifiable

BUT...

- Lot of work
- Lot of skilled experience
- Focused questions
- Multiple experiments
- Context? Complexity?
- Are the measures enough?

EYETRACKING

- Eye-mind hypothesis
- What are people looking at?
 - Fixations
 - Saccades
- Link eye gaze with actions
 - Eg banner blindness



STARBUCKS.COM

[nutrition](#) | [your account](#) | [customer service](#) | [search](#) | [sign in](#)

north america



our coffees



our stores



starbucks card



at home



for business



about us



worldwide

our sites



starbucks store



buy coffee

Make it your drink

Find your favorite drink – hot or cold.

The Ichiro Starbucks Card

A unique and limited-edition Starbucks Card

Starbucks and Ethos

Helping children around the world get clean water.

MAKE IT
YOUR DRINK

search

(keyword or catalog number)

go



Store Locator Find your nearest Starbucks.



We're proud to be on the list We're proud to be among FORTUNE's 100 Best Places to Work.

Subscribe Receive the latest from Starbucks.com.

enter email address

go

Career Center Find your future at Starbucks.

Reload your Starbucks Card It will be ready when you are.

NOT SO SIMPLE

- We skip yet see
- We fixate but don't see
- Drown in data

DESIGN

- Does this work?
- How might it be better?
- Can we improve it quickly?

RESEARCH

- How do people use it?
- What makes it good/bad?
- Does my theory predict usage?

UCD

- Reqs: conceptual models, interviews
- Design: personas, scenarios, claims
- Prototype: sketches, cheap interfaces, visual design
- Evaluate: walkthroughs, heuristics, lab studies

SUMMING UP

- UCD is an attitude
 - Not necessarily at odds with SE...
- Narratives are key
- Exciting time to be an interaction designer

WHY BOTHER?

- Tied to the usability engineering lifecycle
- Pre-design
 - investing in new expensive system requires proof of viability
- Initial design stages
 - develop and evaluate initial design ideas with the user

WHY BOTHER?

- Iterative design
 - does system behavior match the user's task requirements?
 - are there specific problems with the design?
 - what solutions work?
- Acceptance testing
 - verify that system meets expected user performance criteria
 - 80% of 1st time customers will take 1-3 minutes to withdraw \$50 from the automatic teller

USABILITY ENGINEERING APPROACH

- Observe people using systems in simulated settings
 - people brought in to artificial setting that simulates aspects of real world setting
 - people given specific tasks to do
 - observations / measures made as people do their tasks
 - look for problem areas / successes
 - good for uncovering ‘big effects’
 -



USABILITY ENGINEERING APPROACH

- Observe people using systems in simulated settings
 - people brought in to artificial setting that simulates aspects of real world setting
 - people given specific tasks to do
 - observations / measures made as people do their tasks
 - look for problem areas / successes
 - good for uncovering ‘big effects’
 -



USABILITY ENGINEERING APPROACH

- Is the test result relevant to the usability of real products in real use outside of lab?
- Problems
 - non-typical users tested
 - non-typical tasks
 - different physical environment
 - different social context
 - motivation towards experimenter vs motivation towards boss
- Partial Solution
 - use real users
 - task-centered system design tasks
 - environment similar to real situation



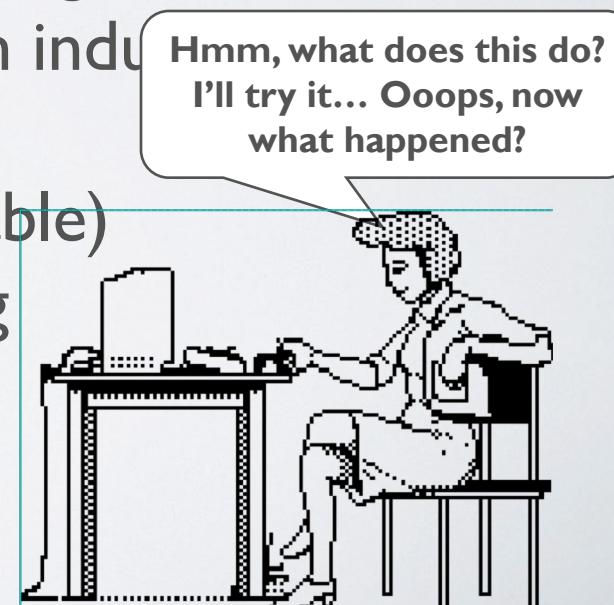
DISCOUNT USABILITY EVALUATION

- Methods
 - inspection
 - extracting the conceptual model
 - direct observation
 - think-aloud
 - constructive interaction
 - query techniques (interviews and questionnaires)
 - continuous evaluation (user feedback and field studies)

Think Aloud Studies

THINK ALOUD METHOD

- Users speak their thoughts while doing the task
 - what they are trying to do
 - why they took an action
 - how they interpret what the system did
- gives insight into what the user is thinking
- most widely used evaluation method in industry
 - may alter the way users do the task
 - unnatural (awkward and uncomfortable)
 - hard to talk if they are concentrating



CODING SHEET EXAMPLE...

Time	General actions			Graph editing			Errors	
	text editing	scrolling	image editing	new node	delete node	modify node	correct error	miss error
09:00								
09:02	X							
09:05							X	
09:10					X			
09:13								

• Tracking a person's use of an editor

- Jakob Nielsen, 1994:
 - Estimating the number of subjects needed for thinking aloud test
- Van Den Haak et al., 2003:
 - Retrospective vs. concurrent think-aloud protocols: testing the usability of a online library catalogue

The think aloud method: A simplified thinking aloud test involves having a test user operate an interface to perform a set of pre-defined tasks while being asked to “think out loud”. By Listening in on the user’s thoughts, the experimenter can pinpoint misconceptions and other usability problems as they occur in the interaction (Nielsen, 1994).

- A lot of developers know about the think-aloud method (21%) but few use it (6%).
- Why ?
 - To complex and expensive
 - The need for highly trained specialists
 - Special equipment and laboratory
- Nielsen however says:
 - There is no need for these expensive and complex tools and specialists if we are only concerned with qualitative results.
 - Because the method can be used at various steps in de design process qualitative results will have a higher impact on the end product.

- two set of experiments using the same think-aloud method.
- The people doing the user testing were two groups of students, three hours of training.
- People doing the tests; 24 and 30 different students.
- Two applications being tested, a commercial word processor and a piece of shareware software.

TABLE 1
Usability problems in a popular commercial word processor

Usability problem	Frequency with which problem was found with one subject
1. Arrow keys do not work even though they are on the keyboard	42%
2. cut/copy/paste hard to learn	42%
3. Changing to a new document hard	33%
4. Scary alert message: change all cannot be undone	29%
5. Two-cursor problem	29%
6. Menu fixation: not seeing the ruler as a place for making formatting changes	25%
7. Confusing menu-based justification commands and ruler-based justification	25%
8. Selected text disappears after insert ruler or insert new page command	25%
9. Hard to distinguish small markers for margin and tabs in the ruler	21%

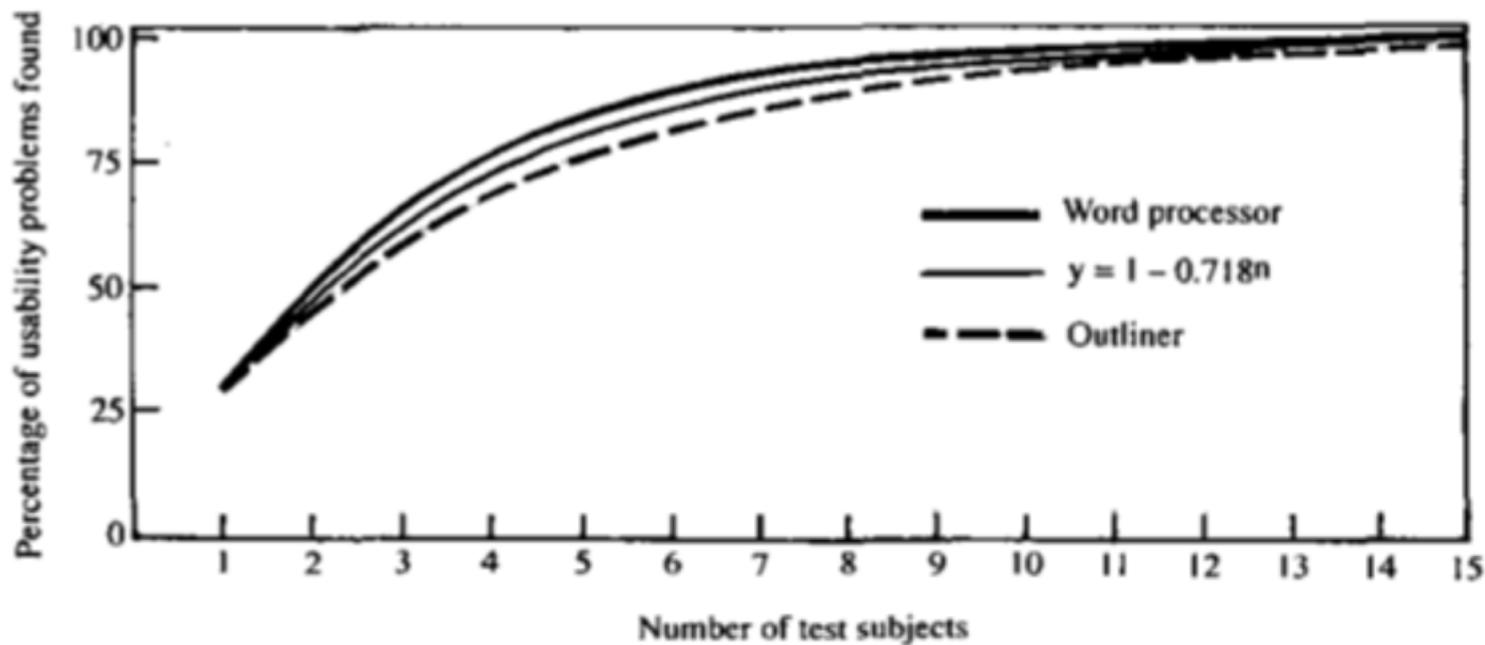


FIGURE 1. The mean percentage of usability problems found by using various numbers of test subjects in a thinking aloud study. The two thick lines show empirically derived numbers and the thin line shows a numerical model.

- The models says that only four to six participants are needed.
- This will give 75% of the usability problems.
- Major problems are more likely to occur during testing than minor ones.

Video prototype

Task 1: try to buy a coffee machine by first creating a user account

Task / Step No	Comments	Issues	Level of Severity
I. User clicks on the logon button	"I can't find the button?" "Where is it?"	Logon Button not visible	critical
2. User tries to register	"ah ok there it is. Ok I need to register first so I go back and click register. The register link is very small"	Link not clearly visible (font size is too small)	major

What & Why of Usability



METHODS



TEMPLATES &
DOCUMENTS



GUIDELINES

How To & Tools



Recruiting Participants & the Legend of
“The General Public”



Content Strategy

Project Management



Basics of User
Experience



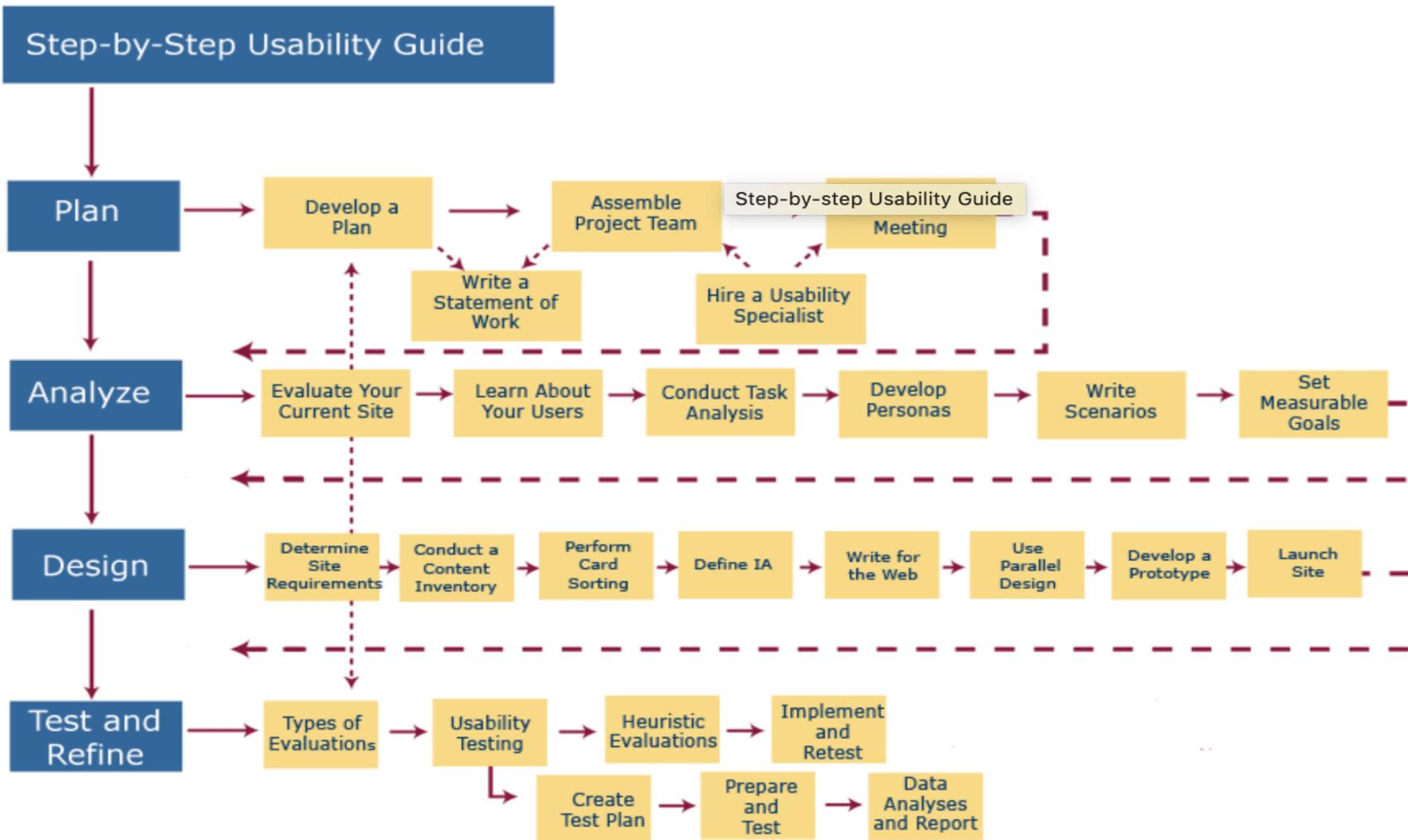
Visual Design



User-Centered Design Process Map

The user-centered design process is [composed of several methods](#) and tasks related to website development. The type of site you are developing, [your requirements, team](#), timeline, and the environment in which you are developing will determine the tasks you perform and the order in which you perform them.

Choose a topic to learn more about how to complete each step.



The System Usability Scale (SUS)

When a SUS is used, participants are asked to score the following 10 items with one of five responses that range from Strongly Agree to Strongly disagree:

- 1.I think that I would like to use this system frequently.
- 2.I found the system unnecessarily complex.
- 3.I thought the system was easy to use.
- 4.I think that I would need the support of a technical person to be able to use this system.
- 5.I found the various functions in this system were well integrated.
- 6.I thought there was too much inconsistency in this system.
- 7.I would imagine that most people would learn to use this system very quickly.
- 8.I found the system very cumbersome to use.
- 9.I felt very confident using the system.
- 10.I needed to learn a lot of things before I could get going with this system.

The System Usability Scale (SUS)

Interpreting scoring can be complex. The participant's scores for each question are converted to a new number, added together and then multiplied by 2.5 to convert the original scores of 0-40 to 0-100. Though the scores are 0-100, these are not percentages and should be considered only in terms of their percentile ranking.

Based on research, a SUS score above a 68 would be considered above average and anything below 68 is below average, however the best way to interpret your results involves “normalizing” the scores to produce a percentile ranking.

ETHICS

- Testing can be a distressing experience
 - pressure to perform, errors inevitable
 - feelings of inadequacy
 - competition with other subjects

Golden rule

- subjects should always be treated with respect

ETHICS – AFTER THE TEST

- Make the users feel comfortable
 - state that the user has helped you find areas of improvement
- Inform the user
 - answer particular questions about the experiment that could have biased the results before
- Maintain privacy
 - never report results in a way that individual users can be identified
 - only show videotapes outside the research group with the user's permission

Ethical Considerations

- .. You must inform participants approximately what they will be asked to do before getting their consent (get them to sign a consent form)
- .. All information is confidential and participants should be told this - you can report results, but not in a way which identifies individuals
- .. Participants should not be subjected to undue stress (tasks that are too difficult to do), bored (evaluations should be interesting and informative) or fatigued (sessions should not be too long)
- .. Participants need to know that they can leave an evaluation at any time if they are unhappy
- .. Participants should be reimbursed appropriately for their time - shouldn't be bribed to act against their better judgements

READING

- Cairns and Cox (2008), *Research Methods in HCI*, chaps 1, 3 and 6
- Chamberlin, Sharp and Maiden (2006) Towards a Framework for Integrating Agile Development and User-Centred Design

READING

- Interaction Design – Preece, Rogers and Sharp
- Chapters 13, 14 and 15
- What do users really care about?: A comparison of usability problems found by users and experts on highly interactive websites – Petrie and Power
- <http://dx.doi.org/10.1145/2207676.2208363>
- The streamlined cognitive walkthrough – Spencer
- <http://dx.doi.org/10.1145/332040.332456>