MMI 2: Mobile Human-Computer Interaction

Evaluation

Prof. Dr. Michael Rohs

michael.rohs@ifi.lmu.de

Mobile Interaction Lab, LMU München

Lectures

#	Date	Topic
1	19.10.2011	Introduction to Mobile Interaction, Mobile Device Platforms
2	26.10.2011	History of Mobile Interaction, Mobile Device Platforms
3	2.11.2011	Mobile Input and Output Technologies
4	9.11.2011	Mobile Input and Output Technologies, Mobile Device Platforms
5	16.11.2011	Mobile Communication
6	23.11.2011	Location and Context
7	30.11.2011	Mobile Interaction Design Process
8	7.12.2011	Mobile Prototyping
9	14.12.2011	Evaluation of Mobile Applications
10	21.12.2011	Visualization and Interaction Techniques for Small Displays
11	11.1.2012	Mobile Devices and Interactive Surfaces
12	18.1.2012	Camera-Based Mobile Interaction
13	25.1.2012	Sensor-Based Mobile Interaction 1
14	1.2.2012	Sensor-Based Mobile Interaction 2
15	8.2.2012	Exam

Review

- What are the pros and cons of iterative design?
- What are the first two questions to answer in the design process?
- What is a “persona”?
- What are scenarios? How can they be represented?
- Strengths and weaknesses of interviews?
- Strengths and weaknesses of questionnaires?
- Strengths and weaknesses of observation?
- The goal of prototyping?

Preview

- From design to evaluation
  - Guidelines
  - Standards
- Measuring usability
  - Usability measures
  - Rating scales for subjective measurements
- Evaluation
  - With users
  - Without users

USABILITY

User – Tool – Task/Goal – Context

context

user

tool

task/goals

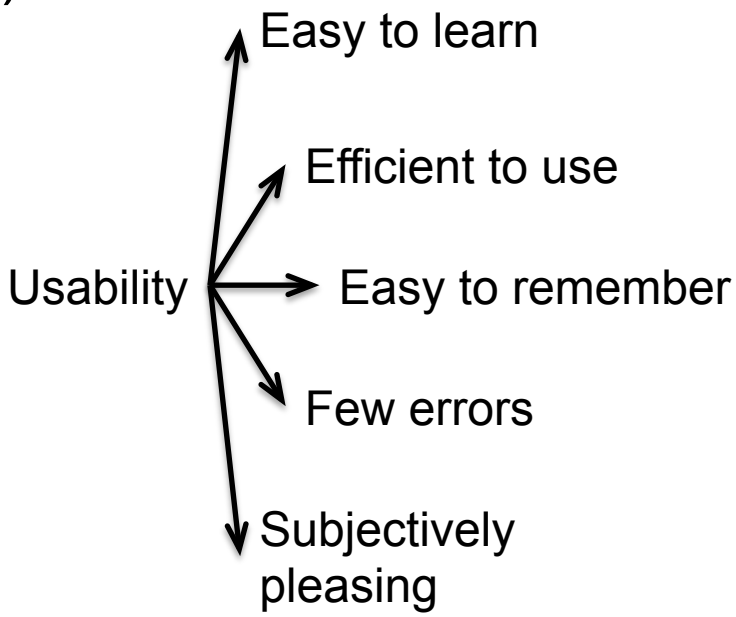
Usability (ISO 9241 Standard)

- Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.
  - Effectiveness: Quality, accuracy, and completeness with which users achieve goals
  - Efficiency: Effort necessary to reach a certain level of quality, accuracy, and completeness
  - Satisfaction: Comfort and acceptability of the system to its users (enjoyable, motivating? or limiting, irritating?)
  - Context of use: Users, tasks, equipment, physical and social environment, organizational requirements

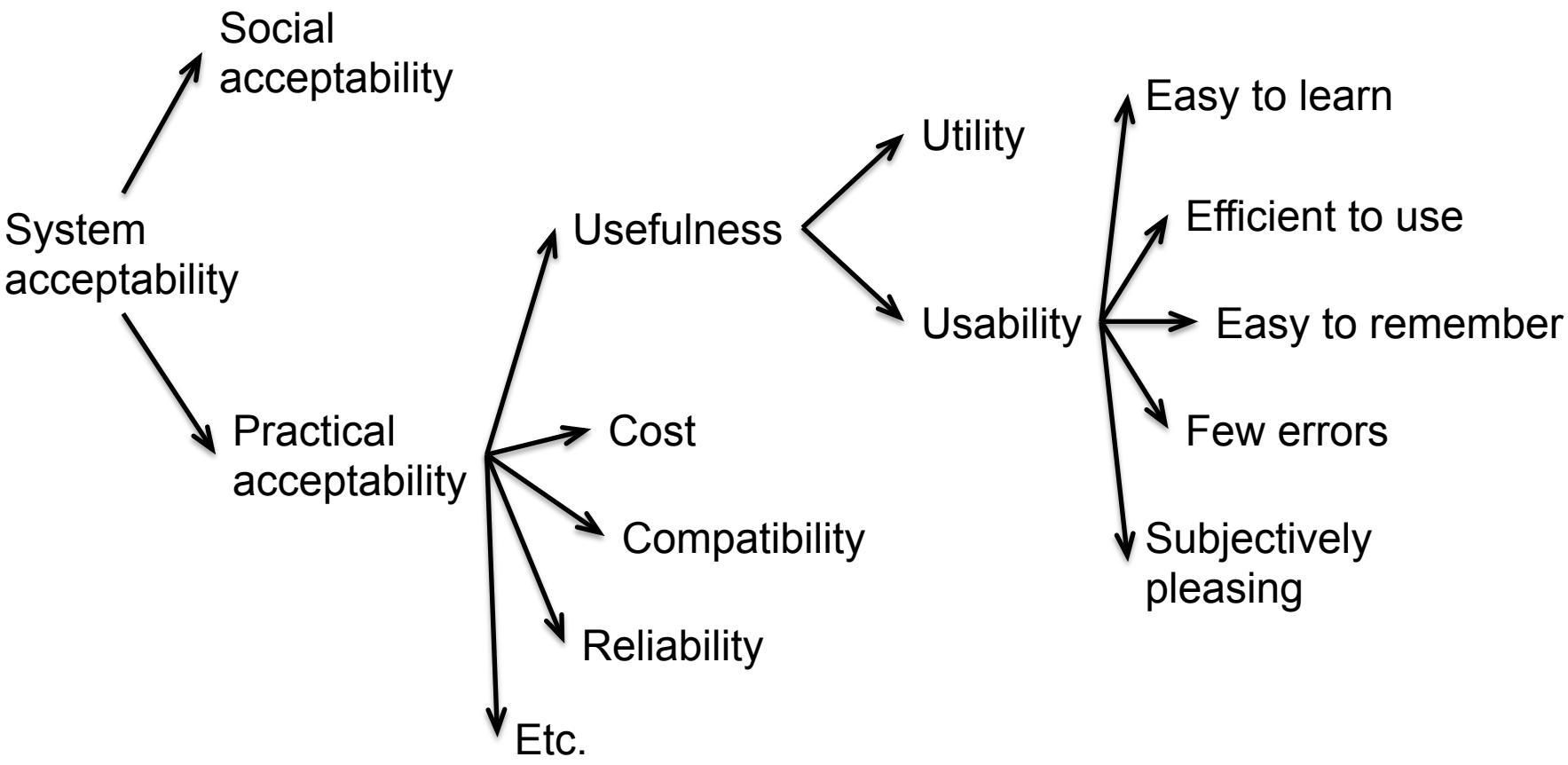
ISO 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11: Guidance on usability—Part 11 (ISO 9241-11:1998)

Attributes of Usability (Nielsen)

- Learnability (easy to learn)
- Efficiency (efficient to use)
- Memorability (easy to remember)
- Errors (few errors)
- Satisfaction (subjectively pleasing)



Usability as an Aspect of System Acceptability (Nielsen)



Typical Measures of Effectiveness

- Binary task completion
- Accuracy
  - Error rates
  - Spatial accuracy
  - Precision
- Recall
- Completeness
- Quality of outcome
  - Understanding
  - Experts' assessment
  - Users' assessment

Kasper Hornbæk: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Human-Computer Studies 64 (2006) 79–102.

Typical Measures of Efficiency

- Time
  - Task completion time
  - Time in mode (e.g., time in help)
  - Time until event (e.g., time to react to warning)
- Input rate (e.g., words per minute, WPM)
- Mental effort (NASA Task Load Index)
- Usage patterns
  - Use frequency (e.g., number of button clicks)
  - Information accessed (e.g., number of Web pages visited)
  - Deviation from optimal solution (e.g. path length)
- Learning (e.g., shorter task time over sessions)

Kasper Hornbæk: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Human-Computer Studies 64 (2006) 79–102.

Typical Measures of Satisfaction

- Standard questionnaires (e.g., QUIS)
- Preference
  - Rate or rank interfaces
  - Behavior in interaction (e.g., observe what users choose)
- Satisfaction with the interface
  - Ease-of-use (e.g. 5-/7-point Likert scale: “X was easy to use”)
  - Satisfaction with specific features
  - Before use (e.g., “I will be able to quickly find pages”)
  - During use (e.g., heart period variability, reflex responses)
- Attitudes and perceptions
  - Attitudes towards others (e.g., “I felt connected to X when using…”)
  - Perception of outcome / interaction

Kasper Hornbæk: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Human-Computer Studies 64 (2006) 79–102.

Typical Measures of Specific Attitudes

- Annoyance
- Anxiety
- Complexity
- Control
- Engagement
- Flexibility
- Fun
- Liking
- Want to use again

Kasper Hornbæk: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Human-Computer Studies 64 (2006) 79–102.

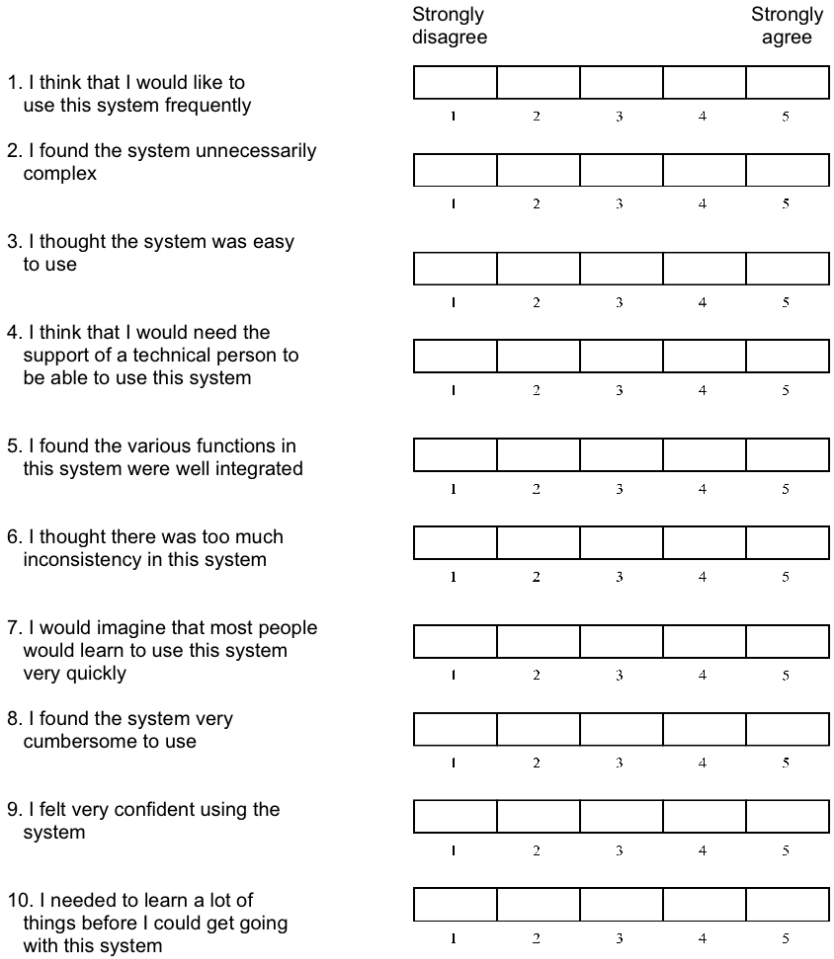
Objective vs. Subjective Measures

- Subjective usability measures
  - Users' perception of attitudes towards interface, interaction, outcome
- Objective usability measures
  - Independent of users' perceptions, physical properties
- Need to study both
  - Subjective may differ from objective measures of time; example: design of progress bars that have shorter subjective time
  - Study found 0.39 correlation between objective and subjective ratings of employee performance

Kasper Hornbæk: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Human-Computer Studies 64 (2006) 79–102.

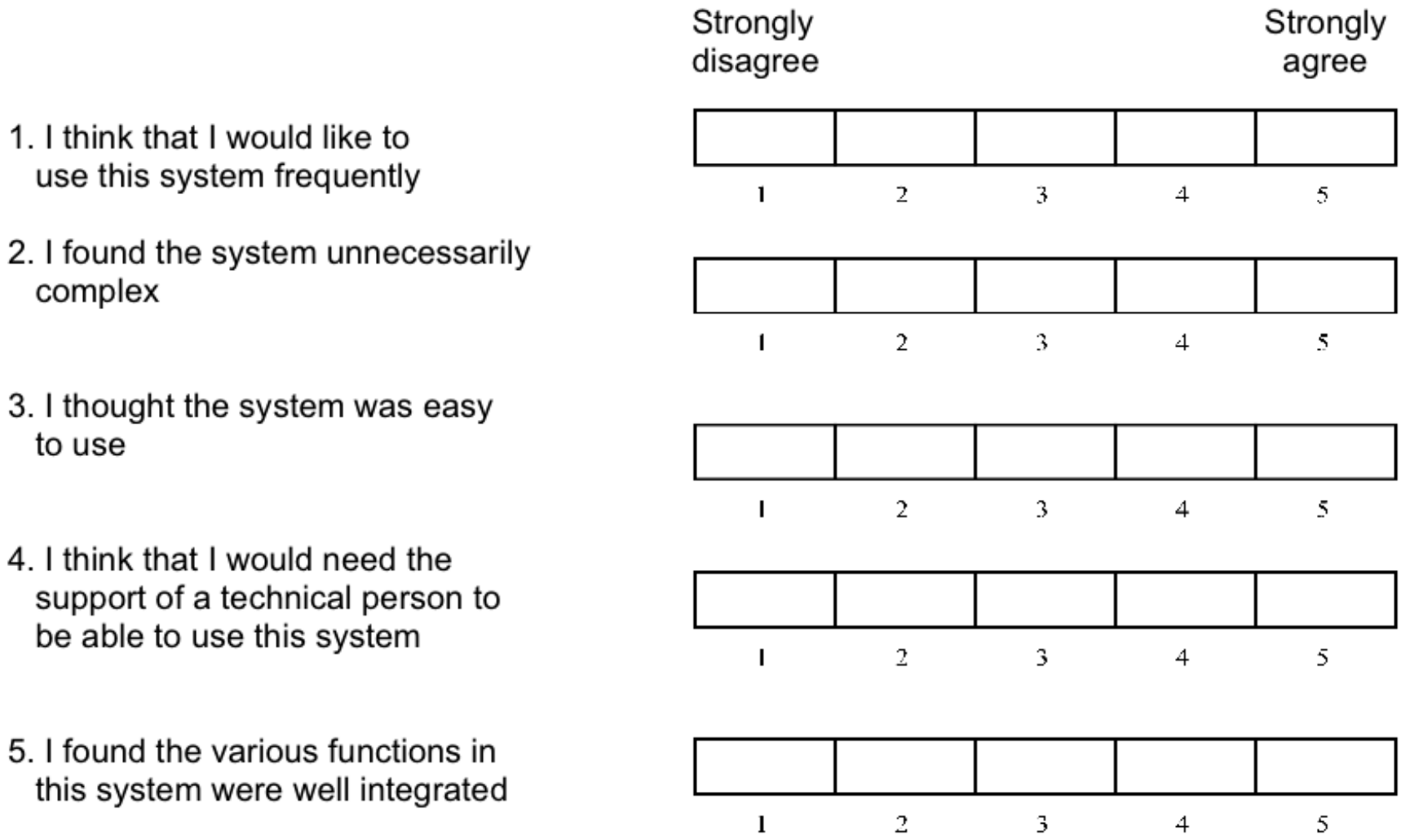
SUS: System Usability Scale

- Developed by DEC Corporation
- 10 5-point Likert scales
- Single score (0-100)
  - Odd items: position – 1
  - Even items: 5 – position
  - Add item scores
  - Multiply by 2.5



Brooke. SUS: A "quick and dirty" usability scale. Usability Evaluation in Industry. London: Taylor and Francis, 1996

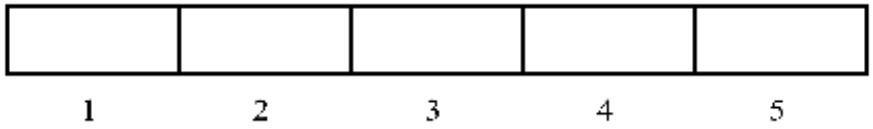
SUS: System Usability Scale



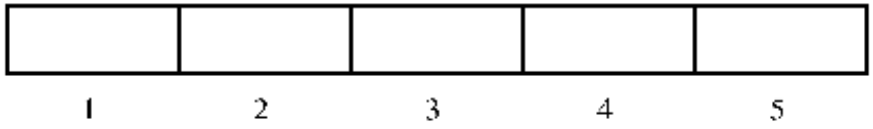
Brooke. SUS: A "quick and dirty" usability scale. Usability Evaluation in Industry. London: Taylor and Francis, 1996

SUS: System Usability Scale

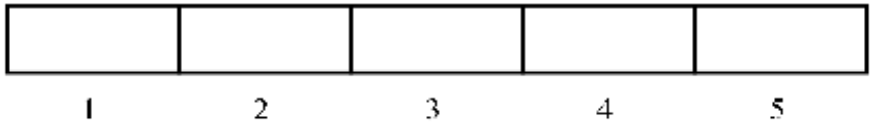
6. I thought there was too much inconsistency in this system



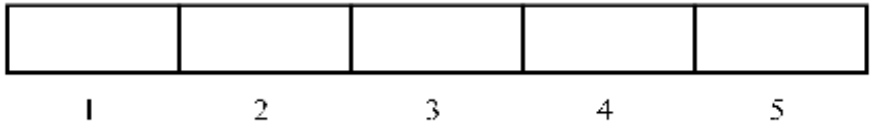
7. I would imagine that most people would learn to use this system very quickly



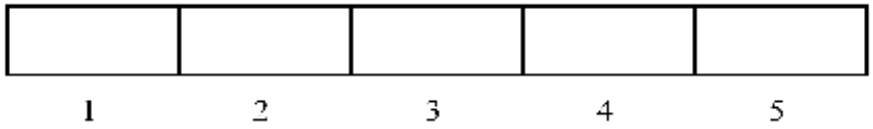
8. I found the system very cumbersome to use



9. I felt very confident using the system



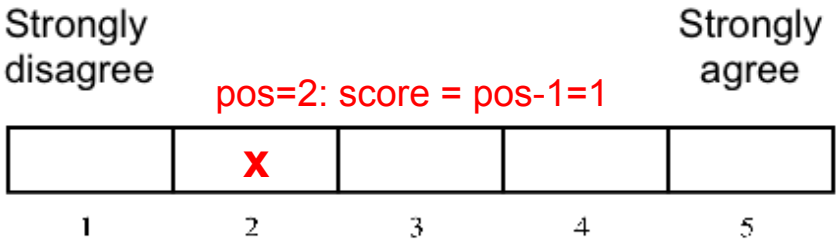
10. I needed to learn a lot of things before I could get going with this system



Brooke. SUS: A "quick and dirty" usability scale. Usability Evaluation in Industry. London: Taylor and Francis, 1996

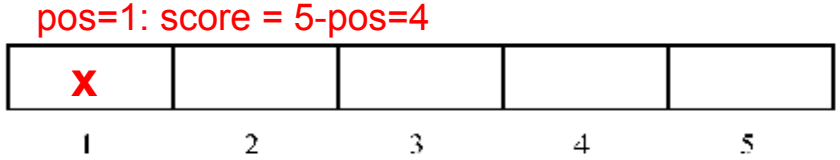
Example: SUS-Ratings

1. I think that I would like to use this system frequently



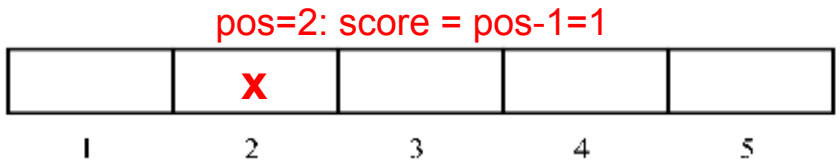
1

2. I found the system unnecessarily complex



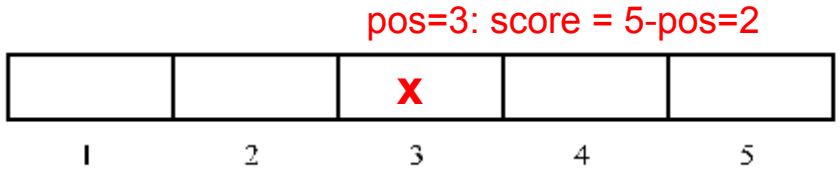
4

3. I thought the system was easy to use



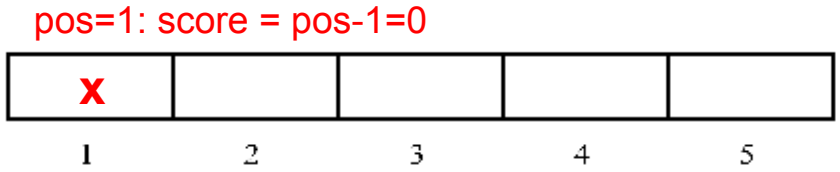
1

4. I think that I would need the support of a technical person to be able to use this system



2

5. I found the various functions in this system were well integrated

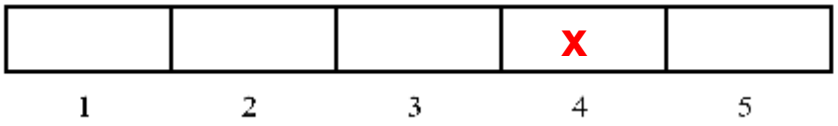


0

Brooke. SUS: A "quick and dirty" usability scale. Usability Evaluation in Industry. London: Taylor and Francis, 1996

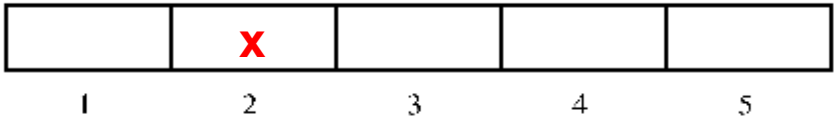
Example: SUS-Ratings

6. I thought there was too much inconsistency in this system



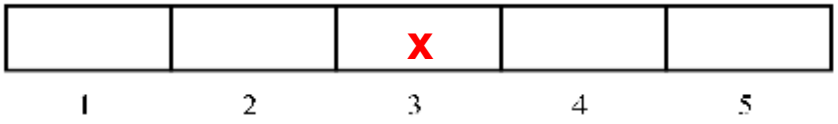
1

7. I would imagine that most people would learn to use this system very quickly



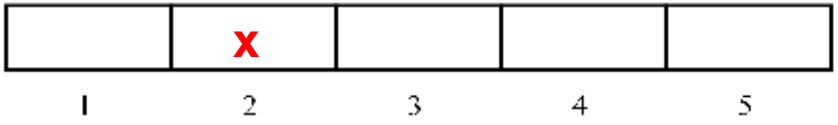
1

8. I found the system very cumbersome to use



2

9. I felt very confident using the system



1

10. I needed to learn a lot of things before I could get going with this system

3

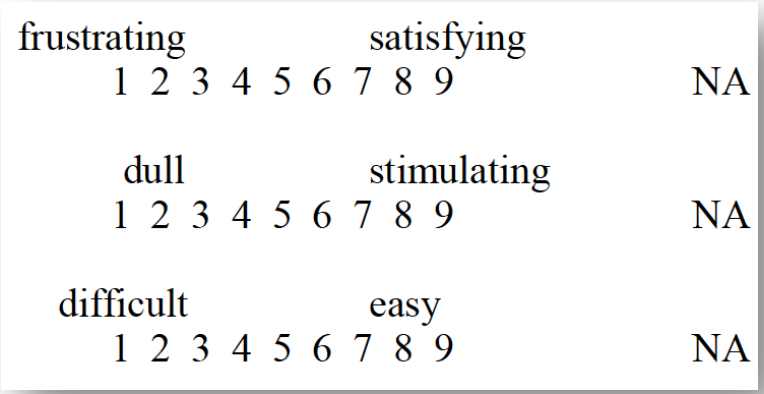
Sum = 16

SUS-Score = Sum * 2.5 = 40

Brooke. SUS: A "quick and dirty" usability scale. Usability Evaluation in Industry. London: Taylor and Francis, 1996

QUIS: Questionnaire for User Interaction Satisfaction

- Developed by the University of Maryland
- Semantic differential scales
- Components: (1) demographics, (2) overall reaction ratings (6 scales), (3) specific interface factors: screen, terminology and system feedback, learning, system capabilities, (4) optional sections
- Long and short forms
- http://lap.umd.edu/quis/

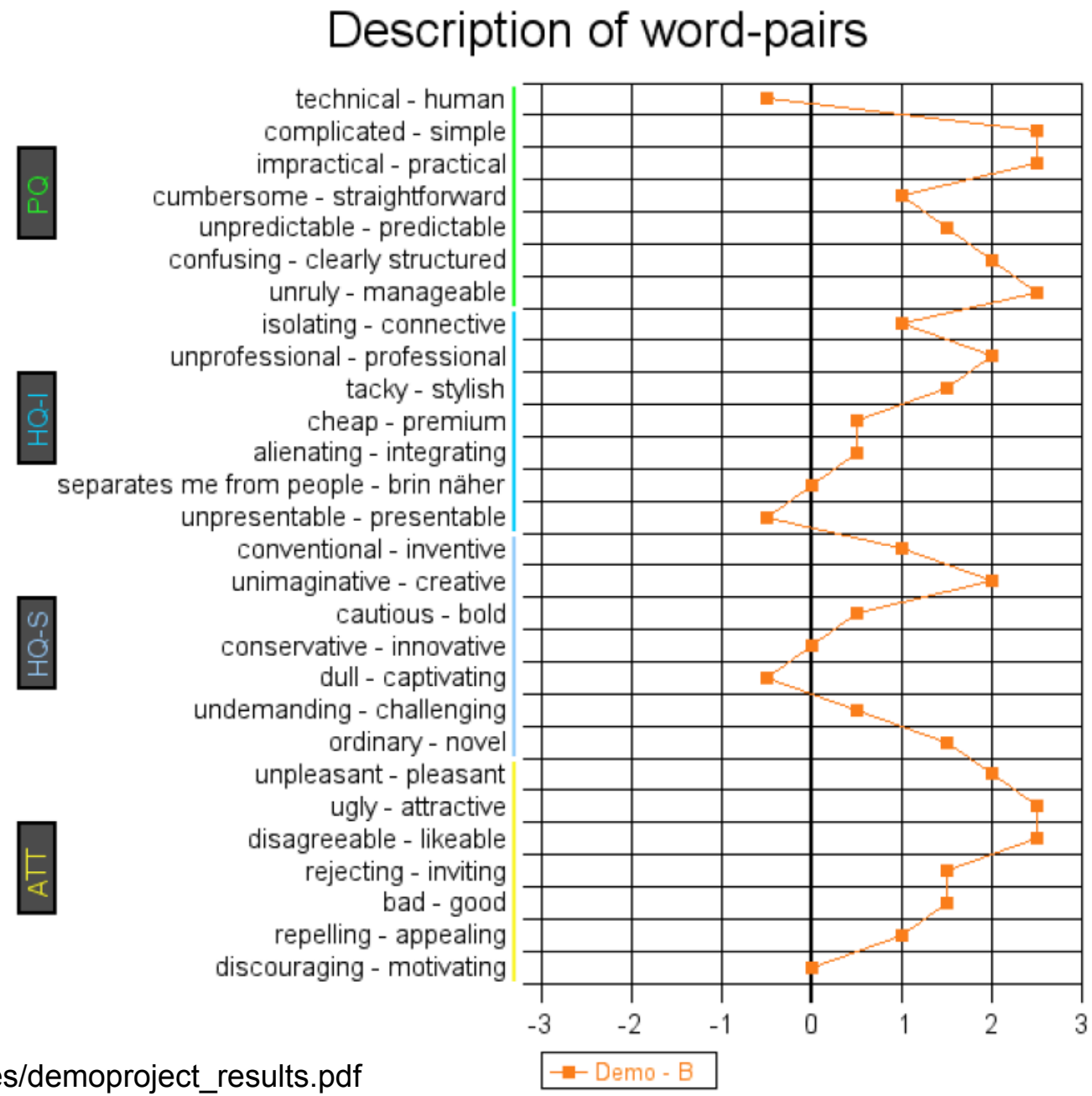


Chin, Diehl, Norman: Development of an instrument measuring user satisfaction of the human-computer interface. CHI '88

AttrakDiff

- Evaluate attractiveness of a product
- Measures pragmatic and hedonic quality
  - Pragmatic quality, e.g., controllable
  - Hedonic quality: identity
  - Hedonic quality: stimulation
  - Attractiveness
- Semantic differential scales
- http://www.attrakdiff.de

AttrakDiff Example



Source:
http://www.attrakdiff.de/files/demoproject_results.pdf

Effect of Prototype Fidelity on Evaluation

- Faster to evaluate sketches instead of functional prototypes
- But: Does representation impact user's perceptions of pragmatic and hedonic quality?
  - Are study results valid?
- Representations
  - Textual description
  - Text & pictures
  - Text & video
  - Text & interaction

candle lamp forget-me-not

Figure 1: Pictures representing the *candle lamp* (left) and *forget-me-not* (right).

Figure 2: Stills from the video representing the *candle lamp* (left) and *forget-me-not* (right).

Diefenbach, Hassenzahl, Eckoldt, Laschke: The impact of concept (re)presentation on users' evaluation and perception. NordiCHI 2010.

Effect of Prototype Fidelity on Evaluation
Study Design and Independent Variables

- N = 326 participants (215 female, ages=15..70)

- 2 factors (independent variables): concept, representation
- 2 levels for concept
  - candle lamp, forget-me-not
- 4 levels for representation
  - text, text & pictures, text & video, text & interaction

- Between-subjects design
  - Each participant randomly assigned to one (concept, representation) pair

Diefenbach, Hassenzahl, Eckoldt, Laschke: The impact of concept (re)presentation on users' evaluation and perception. NordiCHI 2010.

Effect of Prototype Fidelity on Evaluation
Measures / Dependent Variables

- Subjective ratings by users
- Global product evaluation: “Goodness”
  - 7-point semantic differential: bad ←→good
- Perceived product character
  - Pragmatic quality (4 items, e.g., simple ←→ complicated)
  - Hedonic quality (4 items, e.g., dull ←→ captivating)
  - Shortened AttrakDiff2-questionnaire
- Perceived aesthetics of interaction: “Interaction Vocabulary”
  - speed, power, continuity, precision, directedness, spatial proximity, immediacy, change, delay, evidence, need for attention
  - 7-point semantic differential for each item

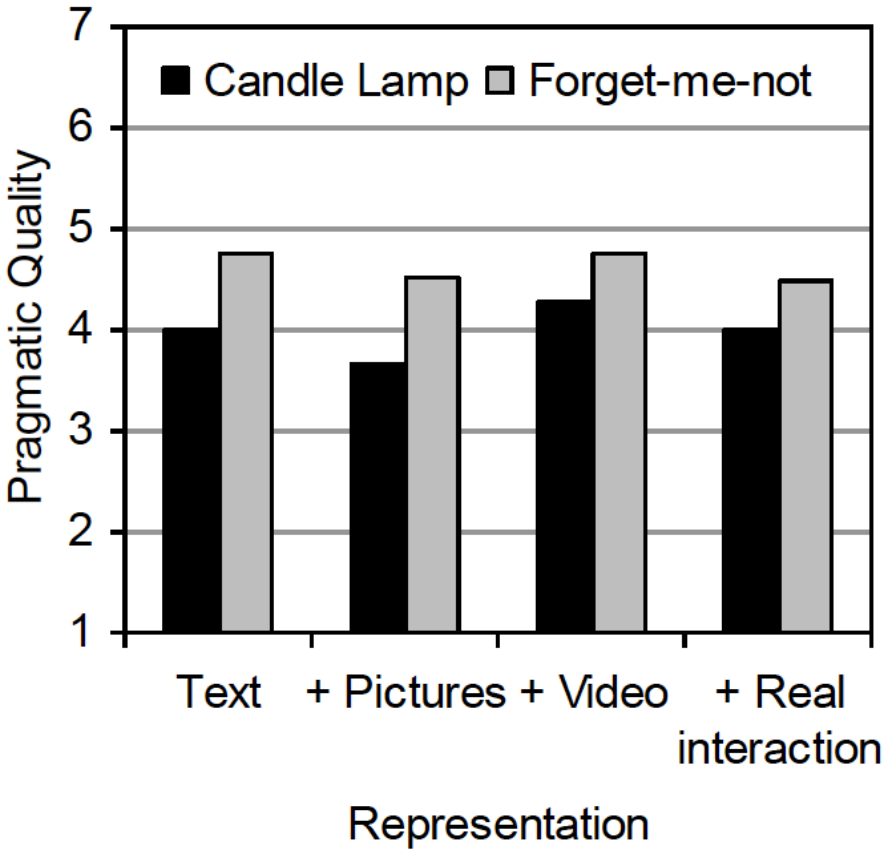
Diefenbach, Hassenzahl, Eckoldt, Laschke: The impact of concept (re)presentation on users' evaluation and perception. NordiCHI 2010.

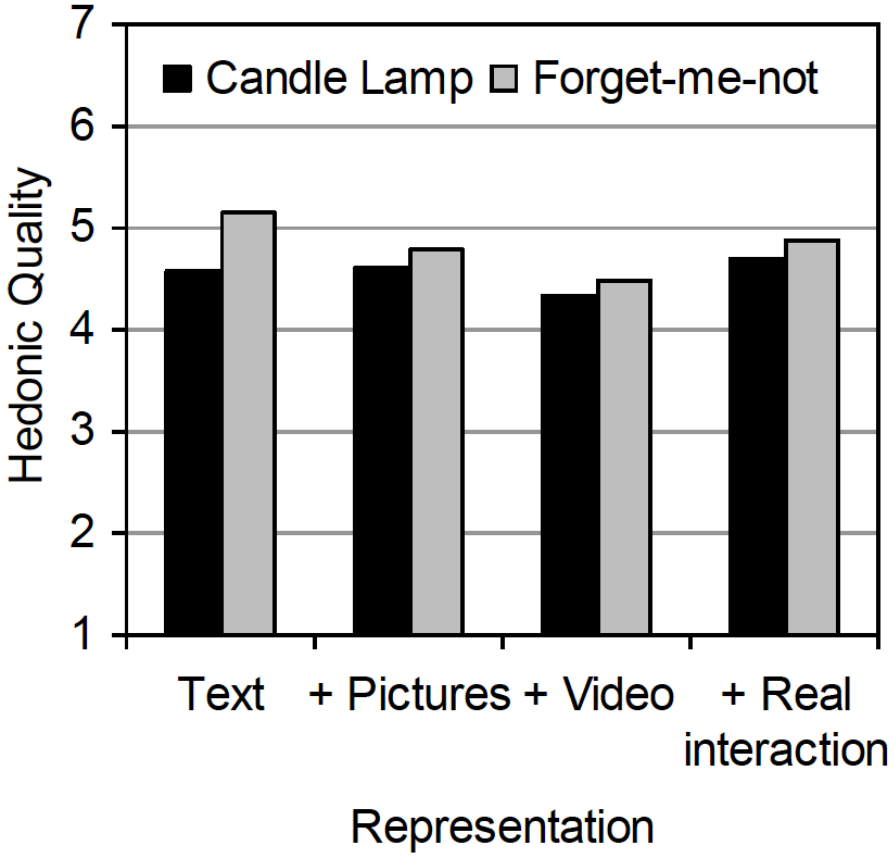
Effect of Prototype Fidelity on Evaluation Results

- No significant effect of representation
  - No impact on global product evaluation (“goodness”)
  - No impact on rating of pragmatic quality
  - No impact on rating of hedonic quality
- Significant effect of system
  - Higher pragmatic quality for forget-me-not
  - Higher hedonic quality for forget-me-not
- Significant effects of representation on perceived aesthetics
  - speed (slow ←→ fast)
  - change (stable ←→ changing)

Diefenbach, Hassenzahl, Eckoldt, Laschke: The impact of concept (re)presentation on users' evaluation and perception. NordiCHI 2010.

Effect of Prototype Fidelity on Evaluation Results





Diefenbach, Hassenzahl, Eckoldt, Laschke: The impact of concept (re)presentation on users' evaluation and perception. NordiCHI 2010.

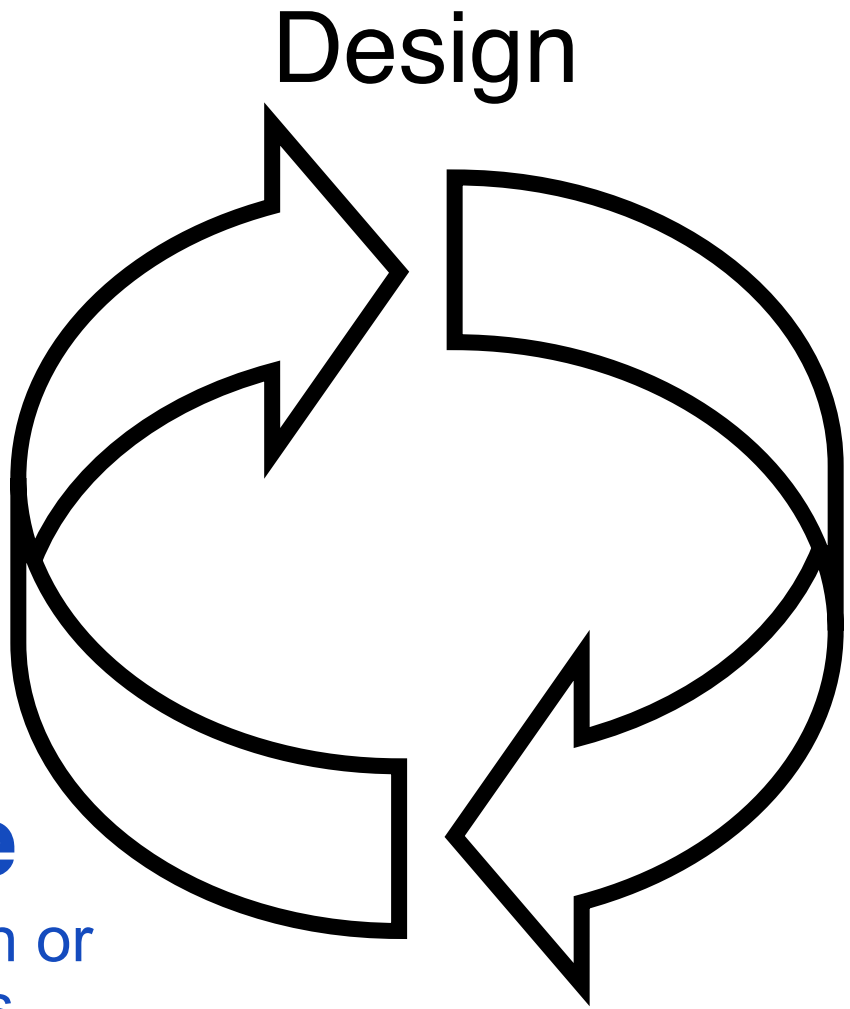
Effect of Prototype Fidelity on Evaluation Results

- Significant effects of representation on perceived aesthetics
  - speed (slow ←→ fast)
  - change (stable ←→ changing)
- Representation has effect on speed / change
- Pairwise comparisons
  - For each pair of representations check whether they yield different speed / change ratings
  - Speed: Interaction faster than Text; Interaction faster than Pictures; Interaction not faster/slower than Video
  - Change: Interaction faster than Picture; no other pairwise effects

Diefenbach, Hassenzahl, Eckoldt, Laschke: The impact of concept (re)presentation on users' evaluation and perception. NordiCHI 2010.

EVALUATION

DIA Cycle: When to evaluate?



Analyze

Evaluate with or without users

Where to evaluate: Laboratory

- With or without users

+ Equipment (audio / video, see-through mirrors, special computers), no disruptions, quiet

– Natural environment missing (shelves, wall calendar, streets, people…); unnatural situation (relevance?)

- Only place possible if real use dangerous, remote (ISS…), or controlled situation needed

Where to evaluate: In the field



- Studies in the users' natural environment
- Advantages
  - + Situations (location and context!) and behavior more natural
  - + More realistic (also because of disruptions)
  - + Better suited to long-term studies
- Disadvantages
  - – Noise, task interruptions
  - – Will still feel like a test situation

Evaluation in the Mobile Context

- Context of use needs to be taken into account
  - Factors: User, activity, device, environment
- Usage “on the move”
  - Physically moving: walking, driving a car, traveling as a passenger
  - Being in different places: away from office environment or home
- Difficult to collect data in the field
  - Recording interaction
  - Capturing context
  - Controlling experimental conditions

Comparison of Lab and Field Tests

- Assess quantity and quality of usability problems found in lab vs. field
- Tasks and scenarios given

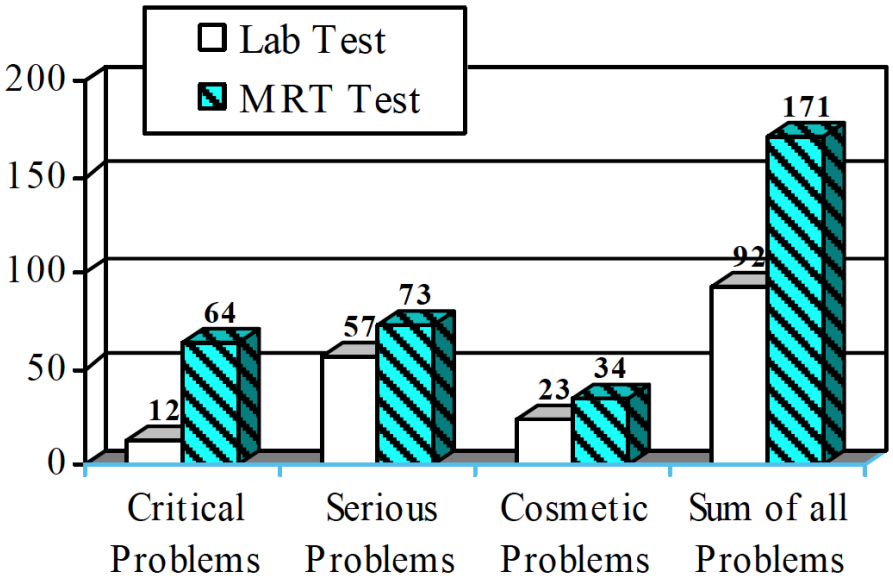
Table 1. The Test Scenario and Tasks involved

Task	Scenario of action	Task Description
1	You need to inform your friend about your personal particulars as he needs to fill up a form for you. You decide to call out.	1. Dial out to contact Gerald from mobile phone contacts list. 2. Start a conversation upon pick up as you normally would. 3. Verbally inform the contact your full name, NRIC, address and date of birth
2	You receive a call from a friend on your mobile phone. You answer the phone call.	1. Answer phone call as you normally would. 2. Start a conversation with the friend.
3	You need to inform your friend about your personal particulars information as he needs to fill up a form for you. You decided to SMS	1. Compose a SMS including the following information: your full name, NRIC, address and date of birth. 2. Send SMS to Gerald from mobile phone contact list. 3. Reply again to Gerald if necessary, i.e. if Gerald replied your message.

Image source: Duh, Tan, Chen: Usability Evaluation for Mobile Device: A Comparison of Laboratory and Field Tests. MobileHCI 2006.

Comparison of Lab and Field Tests

Problems found:



User behavior:

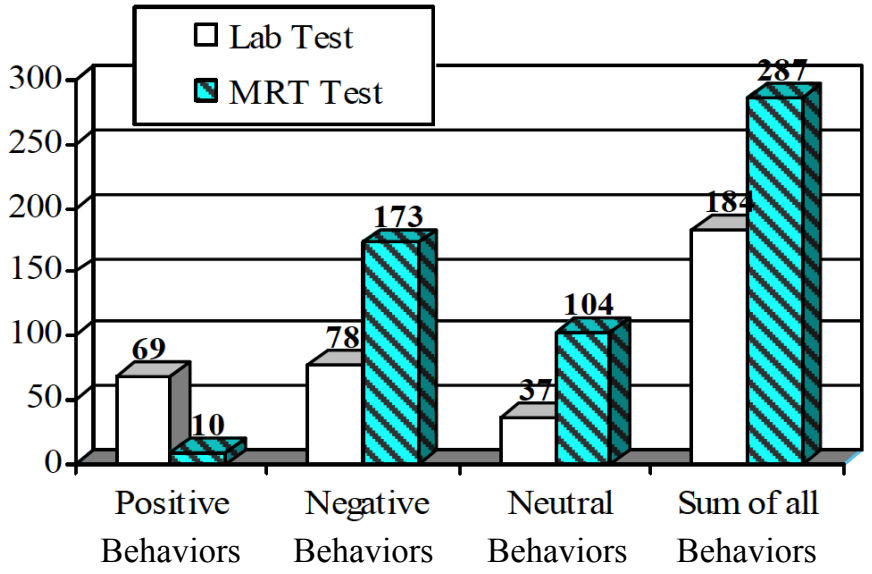


Image sources: Duh, Tan, Chen: Usability Evaluation for Mobile Device: A Comparison of Laboratory and Field Tests. MobileHCI 2006.

Evaluating Attentional Resources in Mobile HCI

- Evaluating the competition for cognitive resources when mobile
- Field study in urban environment
  - Performance of mobile Web tasks
  - Movement through urban situations
- Attention during loading a page
  - Duration of continuous attention
    - Lab: 16.2s → field: 4s
  - Number of attention switches
    - Lab: 1 → field: 8
  - Switching-back durations
    - Railway station: 7-8s, quiet street: 4-6s

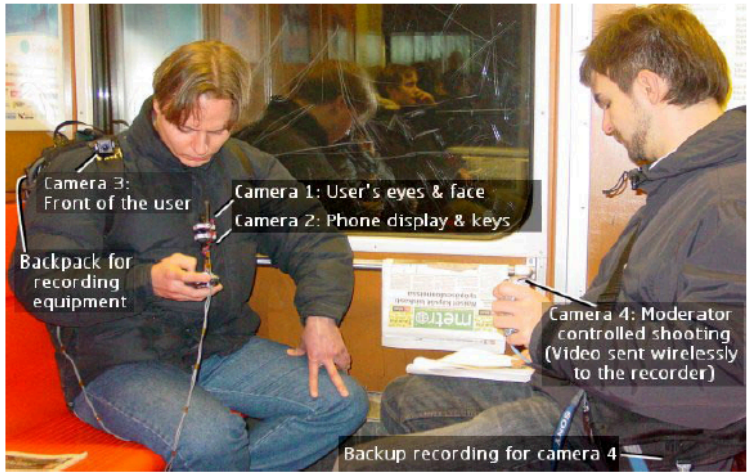


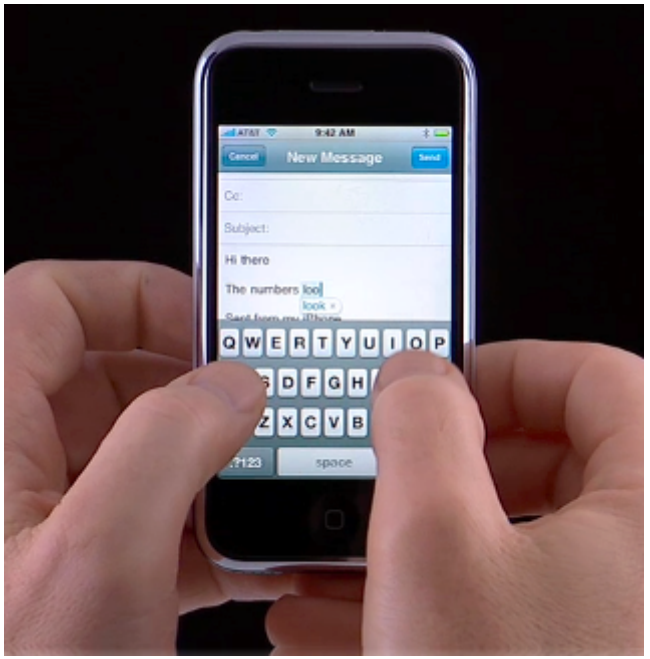
Figure 2. Configuration of recording equipment.

Figure 3. Output video data integrated on-the-fly.

Oulasvirta, Tamminen, Roto, Kuorelahti. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. CHI '05.

Text Input While on the Train

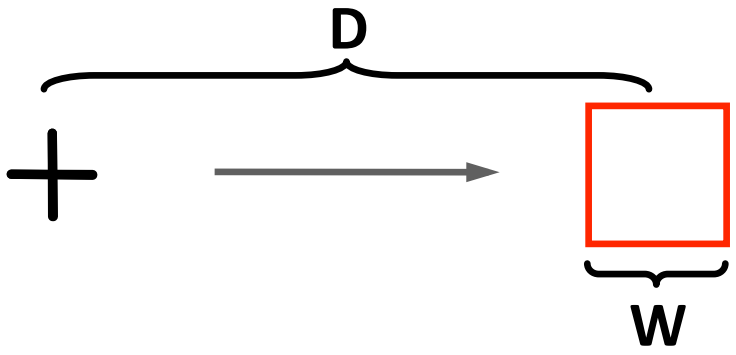
- Touchscreen phones have no tactile feedback for buttons
  - More errors typing text and numbers
- Performance comparison of physical buttons, touchscreen, and touchscreen+tactile
  - In lab and on subway
- Touchscreen+tactile as good as physical buttons
  - Touchscreen alone was poor



Brewster, Chohan, Brown: Tactile feedback for mobile interactions. CHI '07.

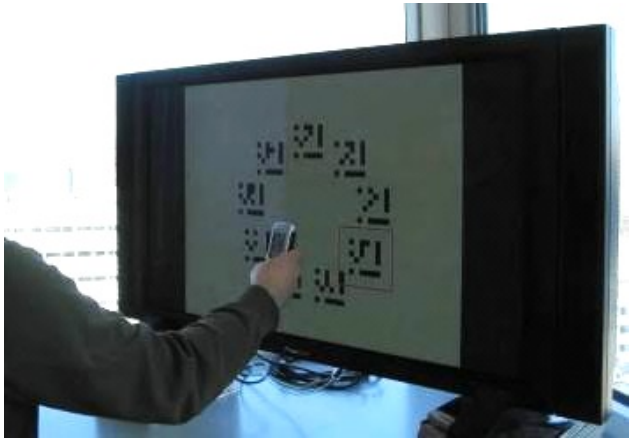
Modeling Mobile AR Pointing with Fitts' Law?

- Goal-directed movement onto target



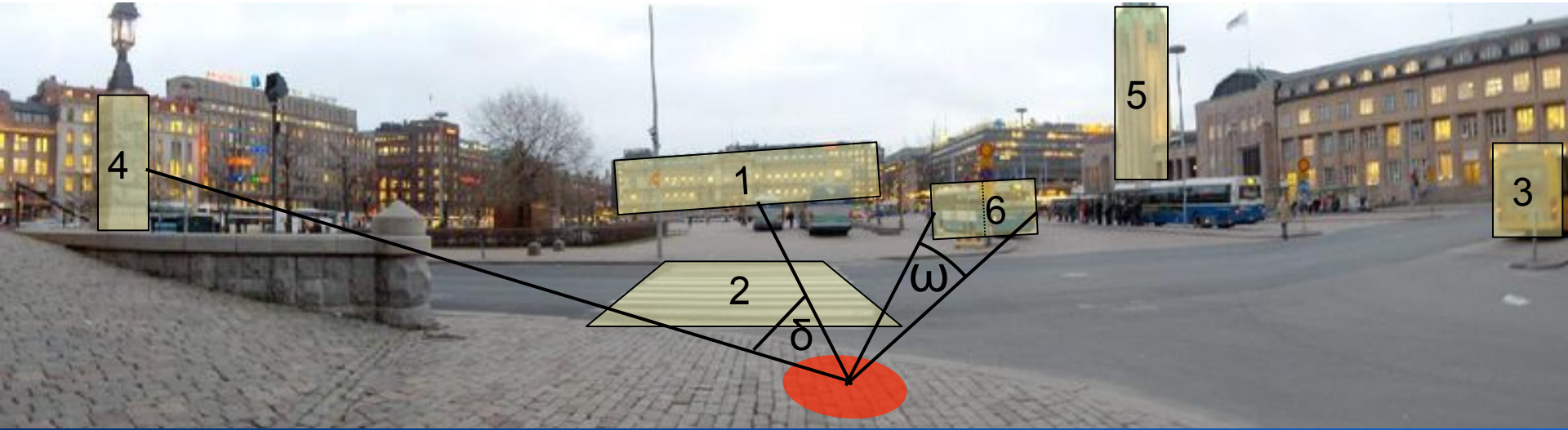
- $MT = a + b \log_2 (D / W + 1)$

- Lab study (Rohs, Oulasvirta, 2008): Fitts' law does not accurately predict movement time for see-through AR pointing

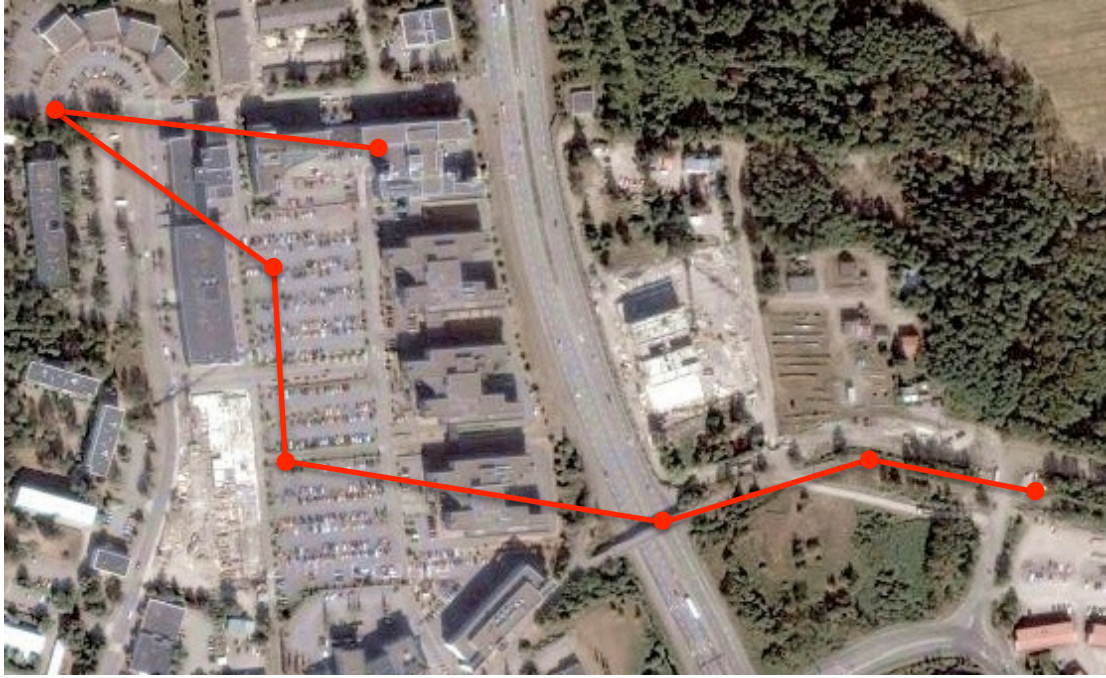


Mobile AR Pointing in the Real World

- 3D targets, varying shape, size, z-distance, visual context
- Angular measure of target distance δ and size ω



Experiment



- 12 participants x 7 sites x 6 target pairs x 24 selections
- Reciprocal pointing task
- ID = 0.72..3.91, D = 6.8°..74.8°, W = 2.3°..35.3°, S = 42.5°
- Saving viewfinder image & timestamp on selection
- Manual post-hoc analysis of selection points

E1: Literature Review

- Many research results about user interface design have been published
- Idea: Search literature for evidence for (or against) aspects of your design

\+ Saves own experiments

– Results only carry over reliably if context (users, assumptions) is very similar

E2: Cognitive Walkthrough

- Analytical method for early design or existing systems
  - Without users
- Expert evaluator = designer or cognitive psychologist
- Goal: Judge learnability and ease of use
  - Does system help user to get from goals to intentions and actions?
- Step through each action and ask
  - Is the effect of the action the same as the user's goal at that point?
  - Will users see that the action is available?
  - Once users find the action, will they know it is the right one?
  - After the action is taken, will users understand the feedback?

E2: Cognitive Walkthrough

- What you need
  - Interface description (prototype of the system)
  - Task description
    - Example: Program the video to time-record a program starting at 18:00 and finishing at 19:15 on channel 4 on January 26, 2011
  - List of interface actions to complete the task
  - User profile
- Doing the actual walkthrough
  - Analyze process of performing the actions using above questions
- Written questions capture psychological knowledge and guide the tester

E3: Heuristic Evaluation

- Choose usability heuristics
  - (general usability principles, e.g., Nielsen's 10 Usability Principles)
- Step through tasks and check whether guidelines are followed
- Severity rating for each problem (Nielsen)
  - 0 = I don't agree this is a problem at all
  - 1 = cosmetic problem
  - 2 = minor usability problem, low priority to fix
  - 3 = major usability problem, high priority to fix
  - 4 = usability catastrophe, imperative to fix before release

\+ Quick and cheap

– Subjective (have several independent evaluators)

See also: www.useit.com/papers/heuristic

10 Usability Principles (Jakob Nielsen)

1. Keep the interface simple!
2. Speak the user's language!
3. Minimize the user's memory load!
4. Be consistent and predictable!
5. Provide feedback!
6. Design clear exits and closed dialogs!
7. Offer shortcuts for experts!
8. Help to recover from errors, offer Undo!
9. Prevent errors!
10. Include help and documentation!

8 Golden Rules of Interface Design (Ben Shneiderman)

1.	Strive for consistency	Sequences, terminology, layout
2.	Cater to universal usability	Diverse users, experience, needs
3.	Offer informative feedback	Direct manipulation, subtle feedback
4.	Design dialogs to yield closure	Grouping of related interactions
5.	Prevent errors	Gray out items, numeric input fields
6.	Permit easy reversal of action	Allow undo, encourage exploration
7.	Support internal locus of control	Minimize surprise, users as initiators rather then responders of actions
8.	Reduce short-term memory load	7 ±2, reduce abbreviation

Rules to Guide the Design Activity

- Follow design rules
  - Restrict space of design options
  - Increase usability of resulting product
  - Judge usability consequences of design decisions
- Classify design rules
  - Authority: Must be followed / just a recommendation
  - Generality: Broadly applicable / very specific situations
  - Level of abstraction
- Rules help to apply theory in practice
  - Design rules based on psychological, cognitive, ergonomic, sociological theory and empirical evidence

Categories of Design Rules

- Principles, Heuristics
  - Small set of general rules (low authority, high generality)
  - Abstract rules, based on psychological knowledge
  - Largely independent of technology
- Guidelines
  - Large set of detailed rules (medium authority, low generality)
  - Often developed for a specific platform
  - More concrete, more technology-oriented
- Standards
  - Agreed upon by a large community (high authority, medium generality)
  - Carefully developed by a standards committee (consensus-based)

Four Fundamental Concepts (Donald Norman)

- Affordances & visibility
  - Affordances
    - http://www.jnd.org/dn.mss/affordances_and_design.html
  - Can the user tell the state of the system and the alternatives for action by looking at the system?
- Conceptual models
  - Is the user able to predict how actions affect the system?
- Natural mapping
  - Is it possible to determine the relationships between actions and results, between controls and effects?
- Feedback
  - Does the user receive full and continuous feedback about the results of actions?

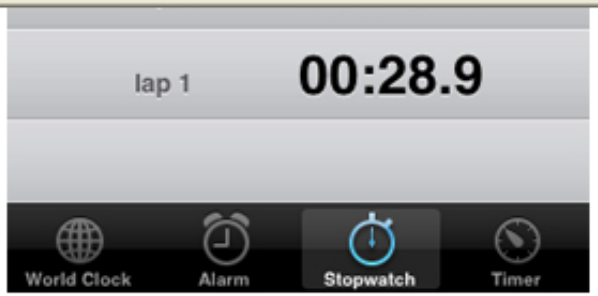
User Interface Guidelines

- Concrete guidelines for look-and-feel and behavior
  - Visual appearance, e.g., icon design
  - Purpose of user interface elements
  - Layout of user interface elements
  - Behavior, conventions of system features
- Android User Interface Guidelines
  - http://developer.android.com/guide/practices/ui_guidelines/index.html
- iOS Human Interface Guidelines
  - http://developer.apple.com/library/ios/documentation/userexperience/conceptual/mobilehig/MobileHIG.pdf
  - Aesthetic integrity, consistency, direct manipulation, feedback, metaphors, user control, …

iPhone Human Interface Guidelines: Simplicity and Ease of Use - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://developer.apple.com/documentation/iPhone/Conceptual/iPhoneHIG/iPhoneDesignPrinciples/chapter_4_section_2.html Go



Avoid Clutter

A webpage that is cluttered with many different sizes and styles of elements, different sizes and colors of text, and gratuitous images presents an unpleasant user experience. Viewed in the small iPhone screen, the negative effects of clutter are magnified, making webpages that might be acceptable on the desktop difficult to use on iPhone.

In both webpages and iPhone applications, it's important to avoid overloading users with a profusion of images and elements. Space is at a premium in the iPhone screen, so you should display only those elements that provide essential information or functionality in the current context. For the most part, avoid displaying elements and images that are purely decorative.

It's also important to avoid leaving too much blank space around your content. If blank space separates important content, users must pan or scroll past it to reach that content. If a lot of blank space is concentrated around the edges of your webpage, it makes your webpage look poorly laid out. Whether you're designing a webpage or an iPhone application, you should use only enough blank space to make controls easy to tap accurately and to make images and text look uncrowded.

Minimize Required Input

Inputting information takes time and attention, whether users tap your controls or use the iPhone keyboard. If your application requires a lot of user input, either all at once or before users can begin using your

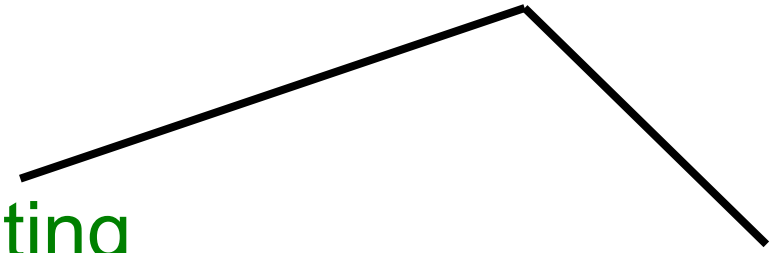
Standards

- Set by national or international standards bodies
  - Ensures acceptance of design rules by a large community
  - Authority through careful design, not always obligatory
- Standards for hardware
  - Ergonomics: understanding human physiology
- Standards for software
  - Psychology: understanding human cognition, motivation, etc.
- Examples:
  - ISO 9241: “Ergonomics of Human System Interaction”, 17 parts
    - 7 parts concerning hardware issues, 8 parts concerning software issues
  - ISO 14915: “Software ergonomics for multimedia user interfaces”, 3 parts
    - “Multimedia navigation and control”, “Media selection and combination”

E4: Model-Based Evaluation

- Several theoretical models exist that offer a framework for design and evaluation
- Examples
  - GOMS (= goals, operators, methods, selection rules)
  - KLM (= keystroke-level model)
  - Design Rationale (history of design decisions with reasons and alternatives)
  - Design Patterns

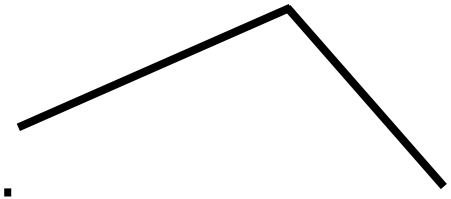
Evaluation Techniques



Evaluating Without Users

E1 Literature Review
E2 Cognitive Walkthrough
E3 Heuristic Evaluation
E4 Model-Based Evaluation

Evaluating With Users



Qualitative

E5 Conceptual Model Extraction
E6 Silent Observation
E7 Think Aloud
E8 Constructive Interaction
E9 Retrospective Testing

Quantitative

E10 Controlled Experiments

+ Interviews, questionnaires,...

Evaluating With Users

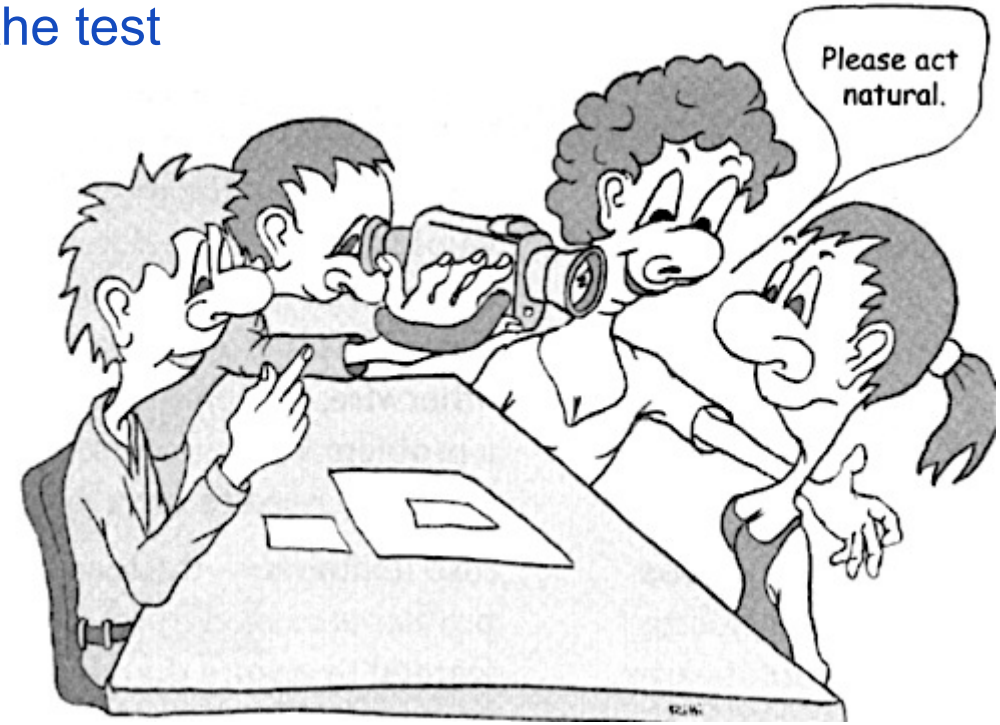
- E1–E4 evaluate designs without the user
- As soon as implementations (prototypes) exist they should also be tested with users, using the following methods

Four Key Issues

1. Setting goals
   - Decide how to analyze data once collected
2. Relationship with participants
   - Clear and professional
   - Protect privacy
   - Informed consent form when appropriate
     - Signed agreement between evaluator and participant
3. Triangulation
   - Use more than one approach
   - Use different perspectives to understand a problem or situation
4. Iterate
   - If questions reveal that goal was not sufficiently refined: refine goal, repeat

Dealing with Test Users

- Tests are uncomfortable for the tester
  - Pressure to perform, mistakes, competitive thinking
- So treat testers with respect at all times!
  - Before, during, and after the test



Data Recording

- Notes, audio, video, photographs

- Notes plus photographs
- Audio plus photographs
- Video

Recording Video in Mobile Evaluation

- Noldus mobile device camera (right)
  - Wireless
- Google setup (left)
  - Observes display and keypad

- Useful if no access to application source code

Schusteritsch, Wei, LaRosa: Towards the perfect infrastructure for usability testing on mobile devices. CHI '07.

www.noldus.com

Mobile Video Capturing Kits

- Composed with common hardware
  - acceptable levels of obtrusiveness
- Monitoring complex scenarios
  - different contexts, users, distractions, etc.
- Synchronizing video + interaction

Figure 4. A user with the shoulder video capturing kit.

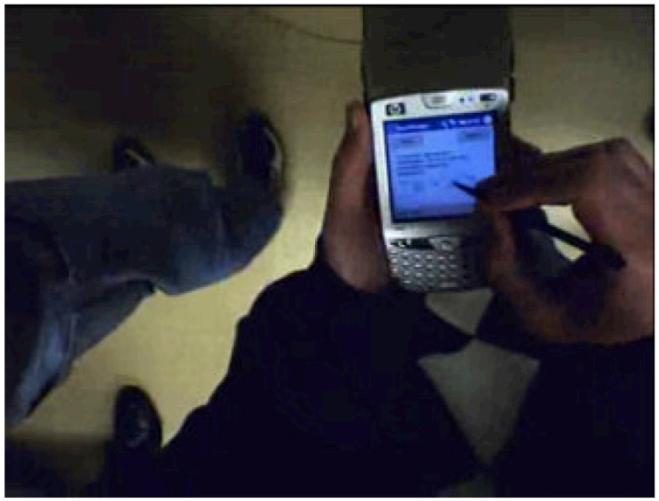


Image sources: de Sa, Carrico: Lessons from early stages design of mobile applications. MobileHCI 2008.

Participatory Design

- Involve user as part of design team throughout entire software process
- Originated in Scandinavia where it is the law
- Techniques for team communication
  - Brainstorming, storyboarding, workshops, interviews, role plays, paper prototypes

E5: Conceptual Model Extraction

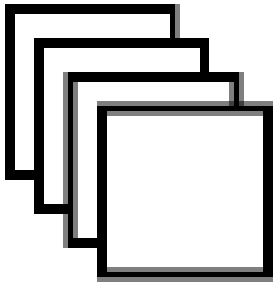
- Designer shows user prototype or screen shots
- User tries to explain elements and their function

\+ Good to understand naïve user's conceptual model of the system

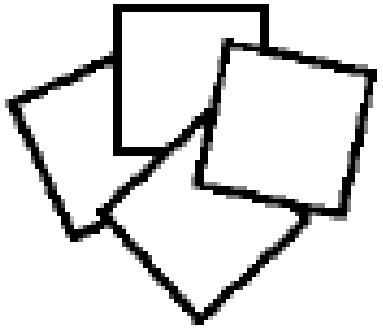
– Bad to understand how the system is learned over time

E5: Conceptual Model Extraction Example

What do these icons mean (in a digital camera)?



Orderly stack



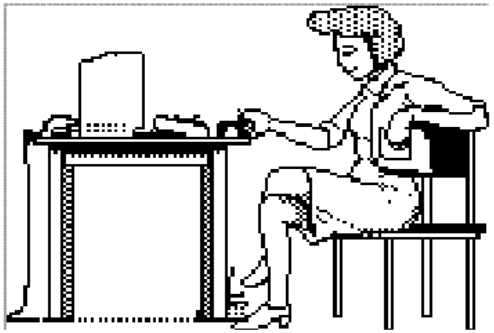
“Messy” stack

Taking pictures of skycrapers?

Viewing taken pictures!

Source: Jones and Marseden: Mobile Interaction Design

Silent Observation



Source: Saul Greenberg

- Designer watches user in lab or in natural environment while working on one of the tasks
- No communication during observation

\+ Helps discover big problems

– No understanding of decision process (that may be wrong) or user's mental model, opinions, or feelings

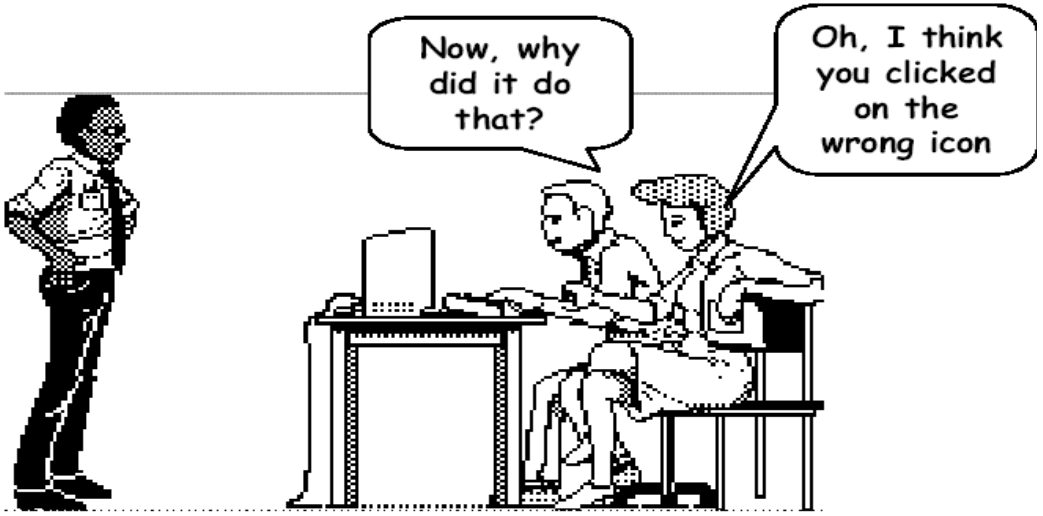
Think Aloud

Source: Saul Greenberg

- As Silent Observation, but user is asked to say aloud
  - What he thinks is happening (state)
  - What he is trying to achieve (goals)
  - Why he is doing something specific (actions)
- Most common method in industry

+ Good to get some insight into user's thinking, but:
  - Talking is hard while focusing on a task
  - Feels weird for most users to talk aloud
  - Conscious talking can change behavior

E8: Constructive Interaction

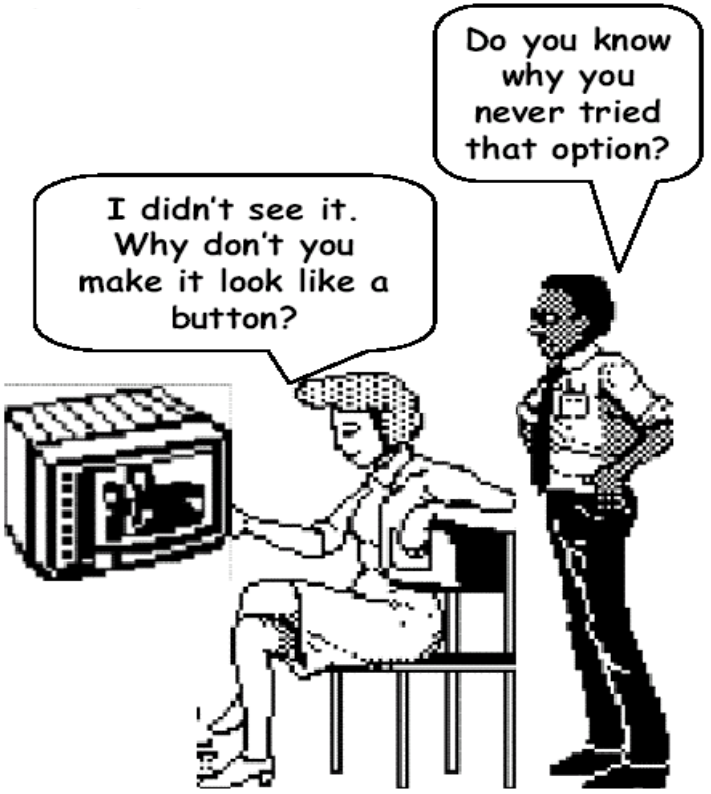


- Two people work on a task together
  - Normal conversation is observed (and recorded)
  - More comfortable than Think Aloud
- Variant of this: Different partners
  - Semi-expert as “trainer”, newbie as “student”
  - Student uses UI and asks, trainer answers
  - Good: Gives insight into mental models of beginner and advanced users at the same time!

Recording Observations

- Paper and pencil
  - Evaluator notes events, interpretations, other observations
  - Cheap but hard with many details (writing is slow)
  - Forms can help
- Audio recording
  - Good for speech with Think Aloud and Constructive Interaction
  - But hard to connect to interface state
- Video
  - Ideal: 2 cameras (user and screen) in 1 picture
  - Best capture, but may be too intrusive initially
- Logging
  - Log input events of the user, synchronize with audio & video

E9: Retrospective Testing



- Additional activity after an observation
- Subject and evaluator look at video recordings together, user comments his actions retrospectively
- Good starting point for subsequent interview, looking at video avoids wrong memories
- Often results in concrete suggestions for improvement

E10: Controlled Experiments

- Quantitative, empirical method
- Steps
  - Formulate hypothesis
  - Design experiment, pick variable and fixed parameters
  - Choose subjects
  - Run experiment
  - Interpret results to accept or reject hypothesis

E10: Controlled Experiments

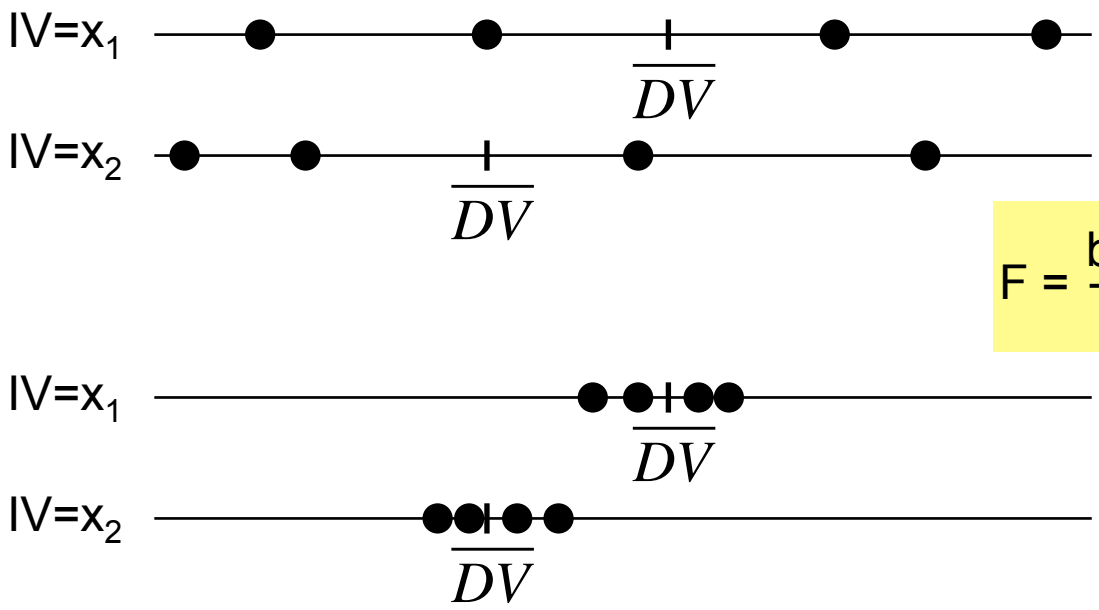
- Subjects
  - Similar to real users in profile
    - Age, education, computer and domain expertise, system knowledge,…
  - Use at least 10 subjects
  - Use more if you need finer details
- Variables
  - Independent: are varied under your control
    - E.g., font size
  - Dependent: are measured
    - E.g., execution time, error rates, subjective preferences

Hypothesis

- A claim that predicts outcome of experiment
  - Example: Reading text in capital letters takes longer than in reading text in small letters
- Hypothesis claims that changing independent variables influences dependent variables
  - Example: Changing small to capital letters (independent variable) influences reading time (dependent variable)
- Experimental goal: Confirm hypothesis
- Approach: Reject null hypothesis (inverse, i.e., “no influence”)
  - Null hypothesis is a term from statistical testing: The samples are drawn from the same statistical distribution

Basic Idea of Statistical Testing

- Assume single independent variable IV with two values
- Take measurements of dependent variable DV for each
- Did IV values have an effect on DV?
  - Assume means are different: due to chance? systematic?
  - How to decide whether there is a relationship?



$$F = \frac{\text{between-groups variability}}{\text{within-group variability}}$$

Basic Idea of Statistical Testing

- Measurements of DV are random samples of populations
- Null hypothesis: all measurements are from one population

  H_0: $\mu_1 = \mu_2$ (population means are equal)
- Alternative hypothesis: not all means are equal
  - Many possibilities, difficult to analyze → focus on H_0
- The larger F, the more likely a systematic effect is present

$$F = \frac{\text{between-groups variability}}{\text{within-group variability}}$$

  - The larger F, the smaller the likelihood of H_0
  - If probability of F is low enough (typically $\alpha = 5\%$): reject H_0 → accept alternative hypothesis

Choosing a Method

- Between-groups
  - – Each subject only does one variant of the experiment
  - – There are at least 2 variants
    (manipulated form & control, to isolate effect of manipulation)
  - \+ No learning effect across variants
  - – But requires more users
- Within-groups
  - – Each subject does all variants of the experiment
  - \+ Less users required, individual differences canceled out
  - – But often learning effect across variants problem

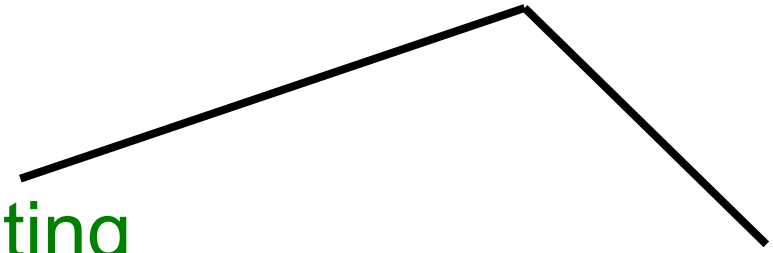
Analyzing Results

- Statistical analysis
  - Often assumptions about underlying distribution
  - t-test: Compare two groups, normal distribution
  - Analysis of variance (ANOVA): Compare two or more groups, normal distribution
  - Regression analysis: How well does result fit to a model?
  - Wilcoxon- or Mann/Whitney test, X^2 test
- Choice depends on
  - Number, continuity, and assumed distribution of dependent variables
  - Desired form of the result (yes/no, size of difference, confidence of estimate)

Other Evaluation Methods

- Before and during the design, with users
  - Personal interviews
  - Questionnaires
- After completing a project
  - Email bug report forms
  - Hotlines
  - Retrospective interviews and questionnaires
  - Field observations (observe running system in real use)

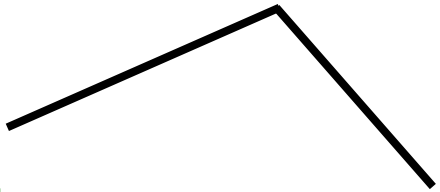
Evaluation Techniques



Evaluating Without Users

E1 Literature Review
E2 Cognitive Walkthrough
E3 Heuristic Evaluation
E4 Model-Based Evaluation

Evaluating With Users



Qualitative

E5 Conceptual Model Extraction
E6 Silent Observation
E7 Think Aloud
E8 Constructive Interaction
E9 Retrospective Testing

+ Interviews, questionnaires,...

Quantitative

E10 Controlled Experiments

Summary

- Evaluate to ensure system matches users' needs
- Evaluation should happen throughout the design process
  - By experts (analytically)
  - By users (experimentally)
- A plethora of methods to evaluate designs
  - Decide when to apply which
- Treat testers with respect at all times!

The End