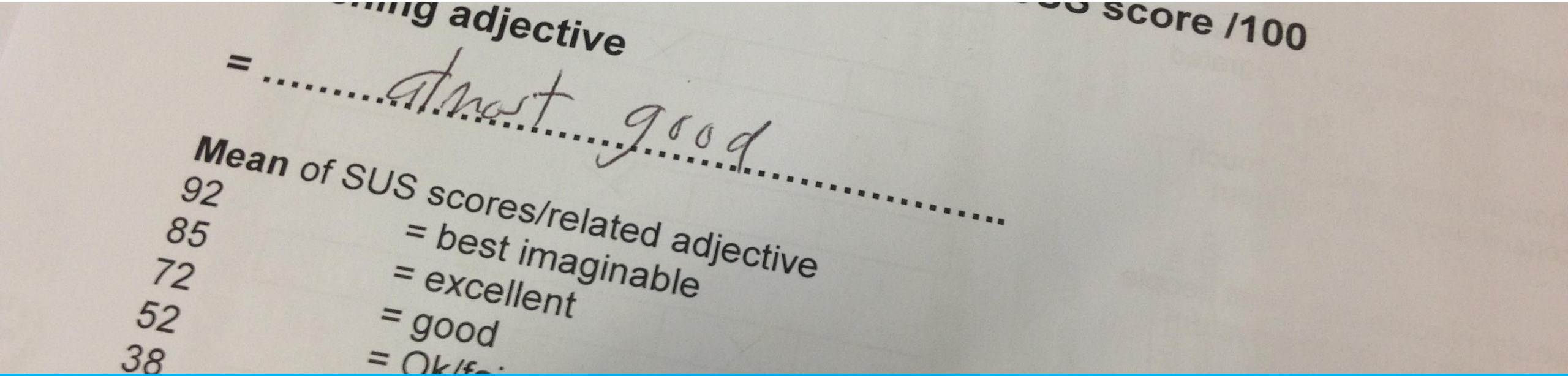


Usability Testing

Image Source: https://pxhere.com/de/photo/205682

Valentin Schwind

Learning Goals

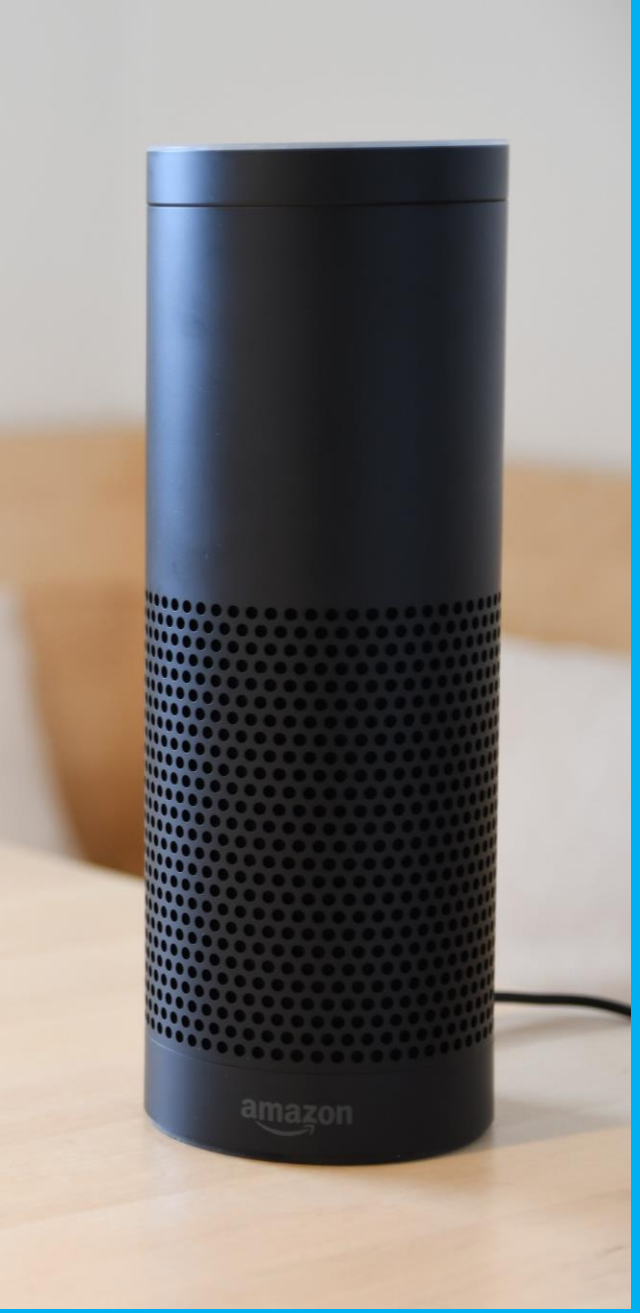
- Understanding the principles of usability testing and user experience sampling
- Understanding when and how usability tests can be conducted
- Learn strengths and weaknesses

Basics

- Testing a prototype with a set of representative users
  - For some tests, users must have certain domain, product and application-specific knowledge and experience
- Users complete typical tasks with:
  - a prototype
  - a part of the system
  - the complete system
  - with competitive systems
- Can be conducted:
  - in early in the design processes to find problems to improve the product (formative evaluation)
  - to validate the design against specific goals (summative evaluation)

Benchmark Testing

- Task-based Usability Tests
  - Ability to complete tasks
  - The time it takes to complete tasks or find information
  - Ability to support users (e.g. learning, reducing errors,…)
- Experienced-based Usability Tests
  - Attitudes toward visual appearance
  - Attitudes toward trust and credibility
  - Perceptions of ease, usefulness, and satisfaction
  - Emotions, immersion, playability (e.g. games)
  - Visibility of an interface

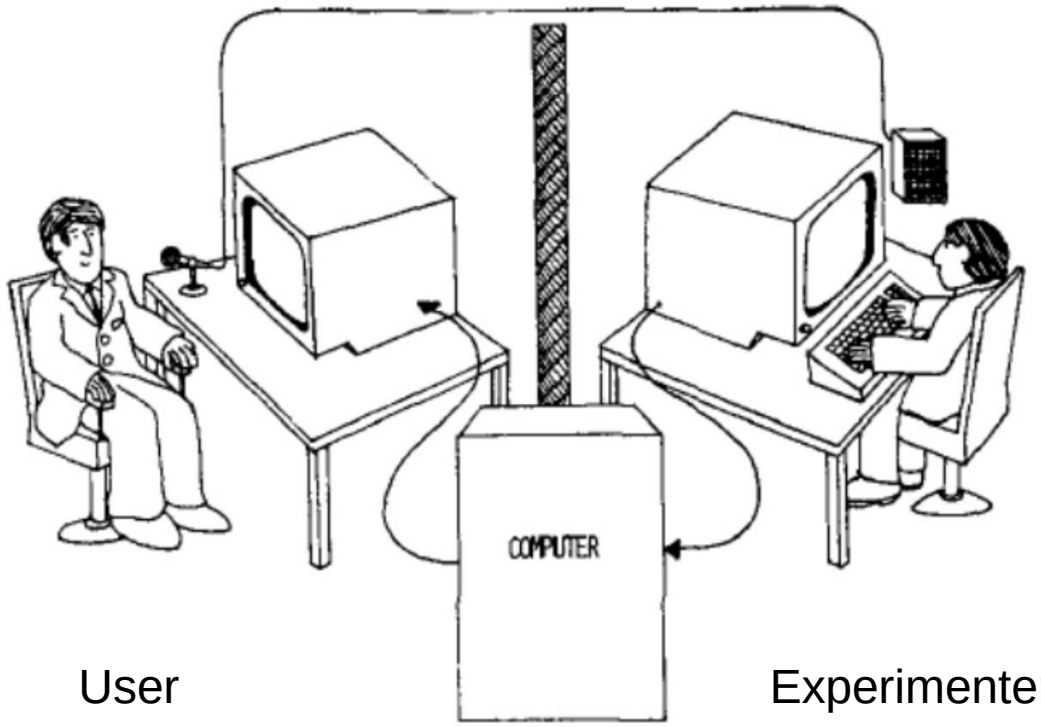


Wizard of Oz

- User-based evaluation of unimplemented technology
- Unknown to the user that a human or team is simulating some or particular functionality of the system
- Examples
  - speech and natural language applications
  - command languages
  - imaging systems
  - pervasive computing applications
  - artificial intelligence

Image Source: https://pxhere.com/cs/photo/1548563

Wizard of Oz Setup



User

Experimenter („Wizard“)

J. F. Kelley. 1983. An empirical methodology for writing user-friendly natural language computer applications. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’83).

Test Locations

Images from: https://pxhere.com ID: 1599048, 718703, 802607, 700764

Interviews

- Unstructured
  - Not directed by a script
  - Rich but not replicable
- Structured
  - Tightly scripted, often like a questionnaire
  - Replicable but may lack richness
- Semi-structured
  - Guided by a script but free to explore interesting issues in more depth
  - Good balance between richness and replicability

How to Ask Questions

- Clear and simple, not too broad
  - “How do you like the UI?” is too general!
- Affording logical, sometimes quantitative answers
  - Bad questions give unusable or wrong answers
  - Open (W-questions) vs. closed questions
- Users don’t always answer truthfully
  - Lack of knowledge, bad estimates, embarrassment
  - So formulate questions carefully, maybe indirectly
- No leading questions!
  - For initial input, do not focus on presenting your design ideas, but on learning about the task

How to Ask Questions

- Clear and simple, not too broad
  - “How do you like the UI?” is too general!
- Affording logical, sometimes quantitative answers
  - Bad questions give unusable or wrong answers
  - Open (W-questions) vs. closed questions
- Users don’t always answer truthfully
  - Lack of knowledge, bad estimates, embarrassment
  - So formulate questions carefully, maybe indirectly
- No leading questions!
  - For initial input, do not focus on presenting your design ideas, but on learning about the task

Running an Interview

1. Introduction: Introduce yourself, explain the goals of the interview, reassure about the ethical issues, ask to record, present any informed consent form
2. Warm-up: Make first questions easy and non-threatening
3. Main body: Present questions in a logical order
4. A cool-off period: Include a few easy questions to defuse tension at the end
5. Closure: Thank interviewee, signal the end, e.g., switch recorder off

- Enriching the Interview: Present prototypes or devices for prompting interviewee
- Data Recording: Notes, audio, video, photographs

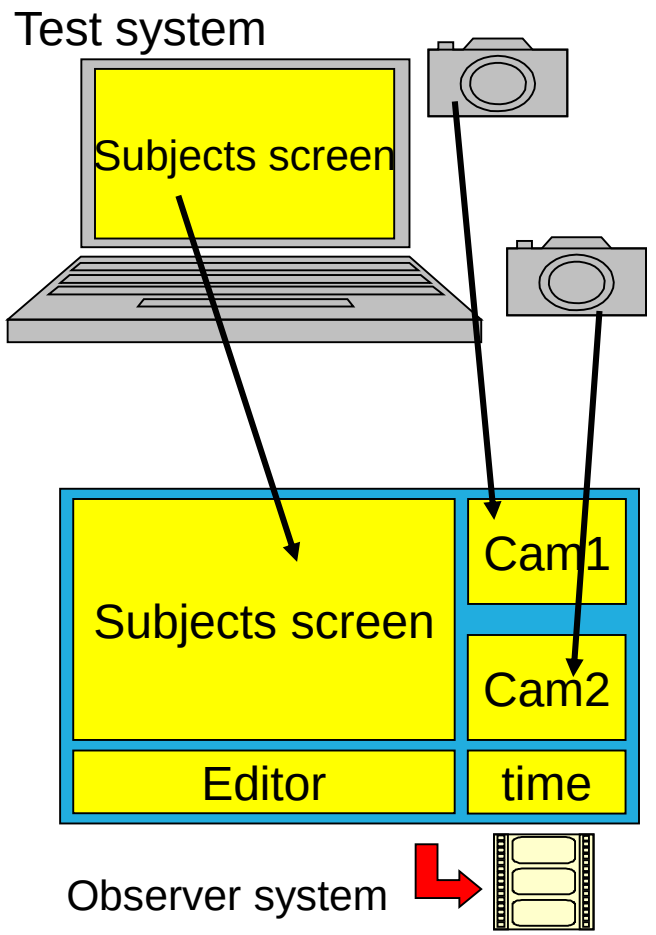
Think Aloud Protocol

- Users are asked to tell what they are
  - doing
  - looking at
  - wondering about
  - feeling
  - thinking
- ….while performing specific tasks during a Usability Test
- Experimenter record statements without commenting or answering questions
- Goal: reveal thoughts of the users to be able to relate to why they act the way they do

Live Video Observation

Example

1. Screen, subject, environment…
2. Computer for the test user,
   - run application to test
   - export the screen (e.g. VNC)
3. Computer for the observer
   - see the screen from the subject
   - attach 2 web cams
   - text input for observer notes
   - capture this screen



Observations

- Moderated/Supervised, with the evaluator observing the participant in real time
  - Screen-sharing/webcam: Adobe Connect, GoTomeeting, Zoom, NetMeeting, LiveLook, UserVue, Skupe, WebEx, Glance, Youguu
- Unmoderated/Unsupervised, with the participant working without direct observation or interaction
  - Live sites/apps: UserZoom, RelevantView, WebEffective, Webnographer (qualitative), ClickTale, ClickHeat (quantiative)
  - Testing conceptual artifacts: online card sorting, OptimalSort, WebSort
  - Surveys

General Procedure

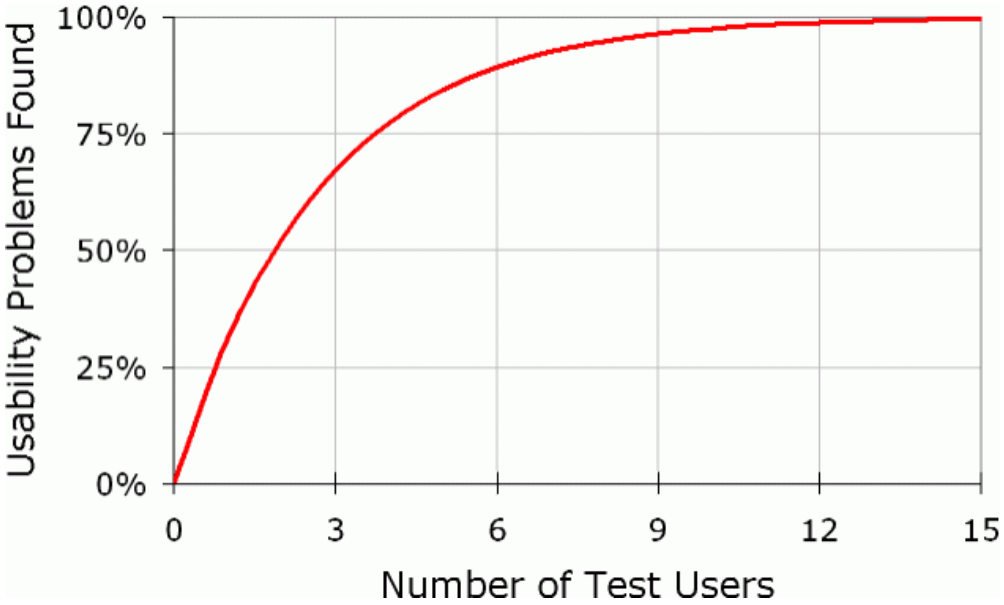
- Test planning
  - Guideline-based, task plan, information material
  - Stakeholder meetings / interviews / participatory design
- Pretest or pilot → inviting participants
- Test sessions
  - Experimenter interviews, watches, listen, take notes
  - Experimenter intervenes when users are clearly in trouble
  - Typically recorded either by video and/or automated testing tool
- Post-test or debrief
- Analysis, interpretation, and presentation of the results.
- Collects ideas for potential improvements

Image from https://www.testingtime.com/blog/usability-test/

Number of Users

- The number of usability problems found in a usability test:

$$ProblemsFound(i) = N\,(1-(1-L)\,n)$$



→ “To find 85% of the usability problems, you need five users.”

https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Feedback Analysis

- Summarize results
  - Transcribe the notes and statements of participants
  - Effectiveness (e.g. task success, number of errors)
  - Efficiency (e.g. task completion time, interaction path, need for assistance)
- Group by usability issue
- Value usability issues
  - Frequency
  - Severity (e.g. caused mistakes, frustration)
- Derive concrete design ideas to improve the prototype
- Questionnaires

Identifying Usability Issues

How important is the task affected by this issue?

The sum of issues for P1+P2+P3

ID	Task	Criticality	Where	Description	Impact	P1	P2	P3	#	Severity
1	Login	3	Login Page	Didn‘t recognize the Facebook option	5	1			0.33	4.95
2	Creating a post	4	Blog	Took too long to load text input page	2		1	1	0.66	5,28
3	Creating a post	5	Blog	Struggled to find the button	3	1	1		0.66	9,9
4	Login	3	Registration	No (auto) login after creating new account	4	1	1	1	1.0	12

Severity = Criticality x Impact x Frequency

Further Approaches

- Contextual Inquiries: researcher watches the user's normal activities and, then, discusses those activities with the user
- Longitudinal Studies
  - Experience Sampling: participants report on their thoughts, feelings, behaviors, and/or environment on multiple occasions over time
  - Diary Studies: participants record entries about their everyday lives in a log, diary or journal about the activity or experiences
- Focus Groups: a heterogonous group of people, whose reactions are studied in guided discussions
  - Brainstorming: generating and evaluating ideas

Further Approaches

- Card sorting/Affinity diagramming: discovering the latent structure in an unsorted list of statements or data
- Context of Use/Scenario of use/Task analysis: collecting and analyzing information about potential users, tasks, and constraints
- Workshops: user and developer identify usability requirements that can be tested in the development process

References

- Dumas, J.S. & Redish, J.C. (1999). A practical guide to usability testing. Exeter: intellect.
- Tullis, T. & Albert, B. (2008). Measuring the user experience. San Francisco: Morgan Kaufmann.
- Joseph S. Dumas (2002). User-based evaluations, The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. Mahwah: Lawrence Erlbaum Associates, Inc.
- David W. Martin (2003). Doing Psychology Experiments. Belmont: Wadsworth Publishing.
- https://measuringu.com/