User Interaction Studies and Usability Testing

Seminar and Discussion

Linnéuniversitetet

2ME30E-VT20

Examensarbete i medieteknik

Today

- Defining an investigation goal
- What is a User Interaction Study?

  Usability, Results, Goals, Tools
- Data collection methods for
  User Interaction Studies
- Conducting a User Interaction Study
- Discussion

  Your thesis proposals and ideas for
  conducting your user interaction study

What (research) questions to ask?

- Before defining concrete (research) questions, try to define the overall aim of your investigation / study in one sentence!

- GQM (Goal/Question/Metric) [1]
  - Purpose
  - Issue
  - Object
  - Viewpoint

via V. R. Basili (1992)

GQM example

- Example: You are interested in visualization approaches for social networks within immersive virtual reality environments.

Purpose: Identify requirements, features and interaction possibilities

Issue: to explore, move and navigate

Object: in dynamically changing social network visualizations

Viewpoint: within immersive virtual reality environments.

What is Usability?

- quality attribute that assesses how easy an user interface (a product) is to use
- 5 components
  - Learnability
  - Efficiency
  - Memorability
  - Errors
  - Satisfaction

via Nielsen Norman Group

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?

- Memorability

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

via Nielsen Norman Group

What is Usability?

- Errors

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

- Satisfaction

  How pleasant is it to use the design?

via Nielsen Norman Group

Terminology

- Utility

  whether it provides the features you need

- Usability

  how easy and pleasant these features are to use

- Useful

  usability + utility

via Nielsen Norman Group

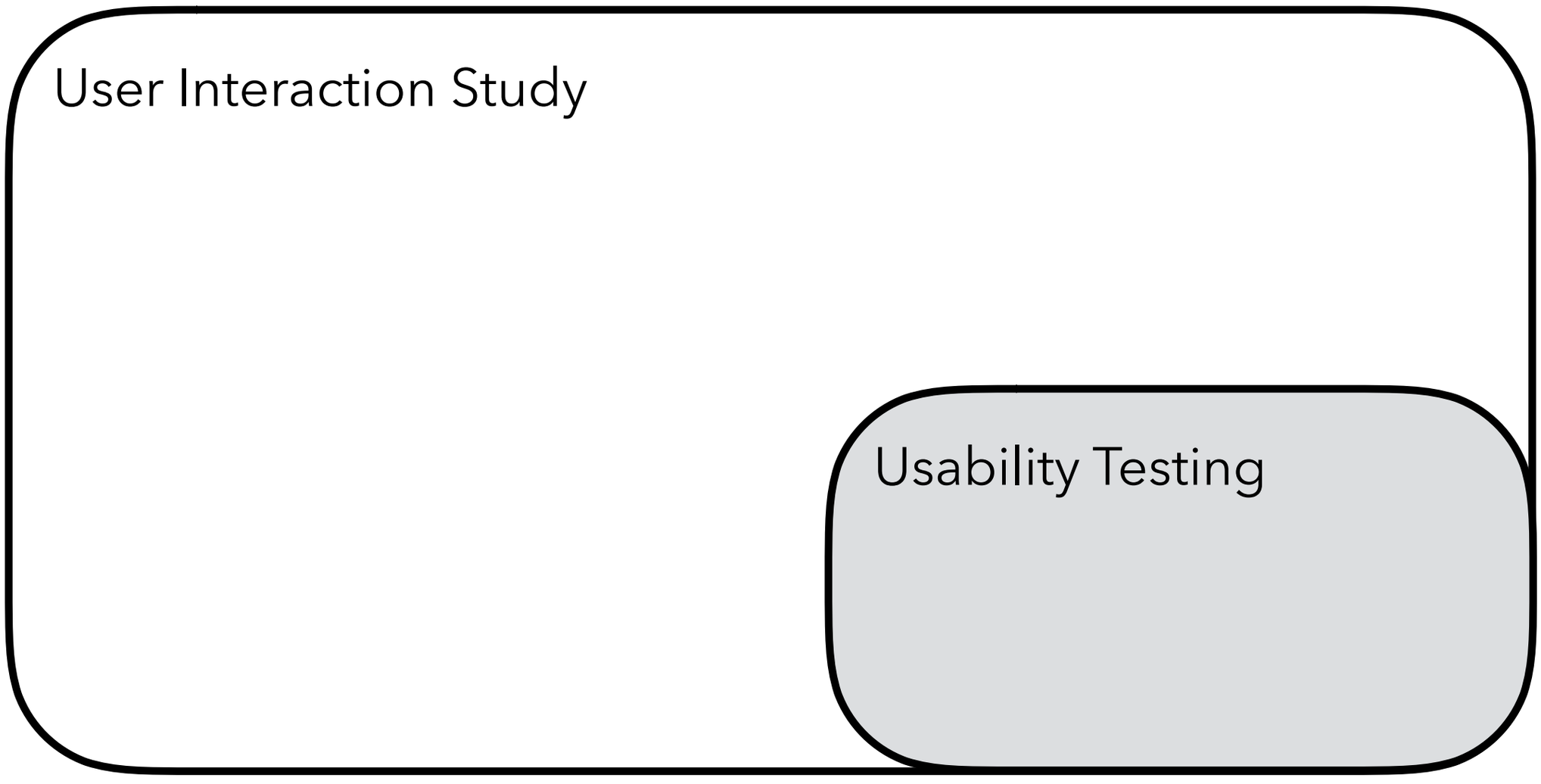
What is Usability Testing?

- process of presenting a product (e.g. developed prototype, website, mobile application…) to a user and ask…
    - to describe what they have in front of them
      (purpose, practical effect, structure, what you can do with it, and so on…)
    - to complete different tasks
      (defined by the researcher)

Usability Testing == User Interaction Study?!

- Usability Testing and User Interaction Studies are **very similar**.
    - User Interaction Study
      The user tests a product (e.g. a software application).
    - Usability Testing
      The user tests a product with the specific goal to investigate its usability.

Usability Testing == User Interaction Study?!



Usability Testing as a sub-group of User Interaction Study.

What is a User Interaction Study? - an example

A user is testing an interactive virtual reality application.

- User has to complete tasks.
- Researcher is observing and taking notes.
- User comments verbally (“thinking-aloud protocol”).
- Video camera records the user interaction.
- Log files record events in the application.

What is a User Interaction Study?

- users of representative target group complete a typical task / typical tasks
- researchers accompany / observe the user interaction study in order to discover new insights,

  e.g. usability problems
- application of common tools to collect data,

  such as Thinking-aloud protocol, Co-discovery learning, Self-constructed questionnaires, System Usability Scale (SUS), User Engagement Scale (UES), AttrakDiff, NASA Task Load Index (TLX), Simulator Sickness Questionnaire (SSQ), Flow Short Scale (FKS), Logging, Explorative Expert Discussion, Conceptual Walkthrough, …

User Interaction Study: Results

- qualitative descriptions of problems

  e.g. “I didn’t find feature X.”

  e.g. “I couldn’t figure out how to do X.”

- quantitative statements

  e.g. “How often…”

  e.g. “How long…”

- subjective assessments

  e.g. “understandable user interface design”

  e.g. “pleasant color palette”

User Interaction Study: Goals

- classic
    - identify advantages and disadvantages in the product design,
      e.g. problems with the usability of the product
    - document identified problems and discovered advantages
      in a report
    - suggest improvements (“re-design”)
- political
    - proof of concept (e.g. interaction design)

Thinking-aloud protocol

- “In a thinking aloud test, you ask test participants to use the system while continuously thinking out loud – that is, simply verbalizing their thoughts as they move through the user interface.”

- Benefits

  cheap, robust, flexible, convincing, easy to learn

- Downsides

  unnatural situation, filtered statements, biasing user behavior

via Nielsen Norman Group

Co-discovery learning

- two users complete a user interaction study together at the same time, e.g. completing tasks, while being observed
- the users can help each other as they were a team in order to accomplish a common goal
- more natural situation than thinking aloud protocol, since users don’t talk to themselves but to each other

via UsabilityHome.com

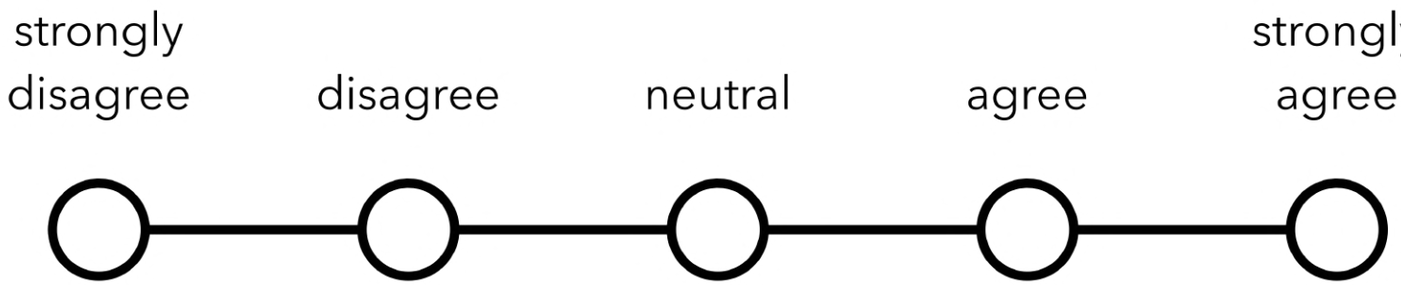
Self-constructed questionnaires

- pre- and post-session (if needed)
- mixture of
    - Likert scale statements (quantitative data)
    - open answer questions (qualitative feedback)

      e.g. “What did you like most / least about …”
- use computerised data collection

  (this will help you a lot with the data analysis!!!)

Likert scale statements



Mean and SD (calculation example)

```
var statement = "I felt interested in this experience.";
var participants = 10;

// coding: strongly disagree (1), disagree (2), neutral (3),
   agree (4), strongly agree (5)

var answers = [5, 3, 4, 2, 4, 4, 5, 2, 5, 4];
```

```
var mean = ?                    // average
var standard_deviation = ? // square root of variance


// helper values (in order to calculate standard deviation)
var d1, d2, ..., dn = ?    // single squared deviations from mean
var variance = ?           // mean of all single squared deviations
```

Mean and SD (calculation example)

```
var answers = [5, 3, 4, 2, 4, 4, 5, 2, 5, 4]; // n = 10

// pseudo code: calculations
// average
var mean = (sum of all answers) / count of answers
```

```
// single squared deviations from mean
var d1 = squared (first answer - mean)
var d2 = squared (second answer - mean)
...
var dn = squared (n_th answer - mean)


// mean of all single squared deviations
var variance = sum of all single squared deviations from mean /
count of answers


// standard deviation
var standard_deviation = square root of variance
```

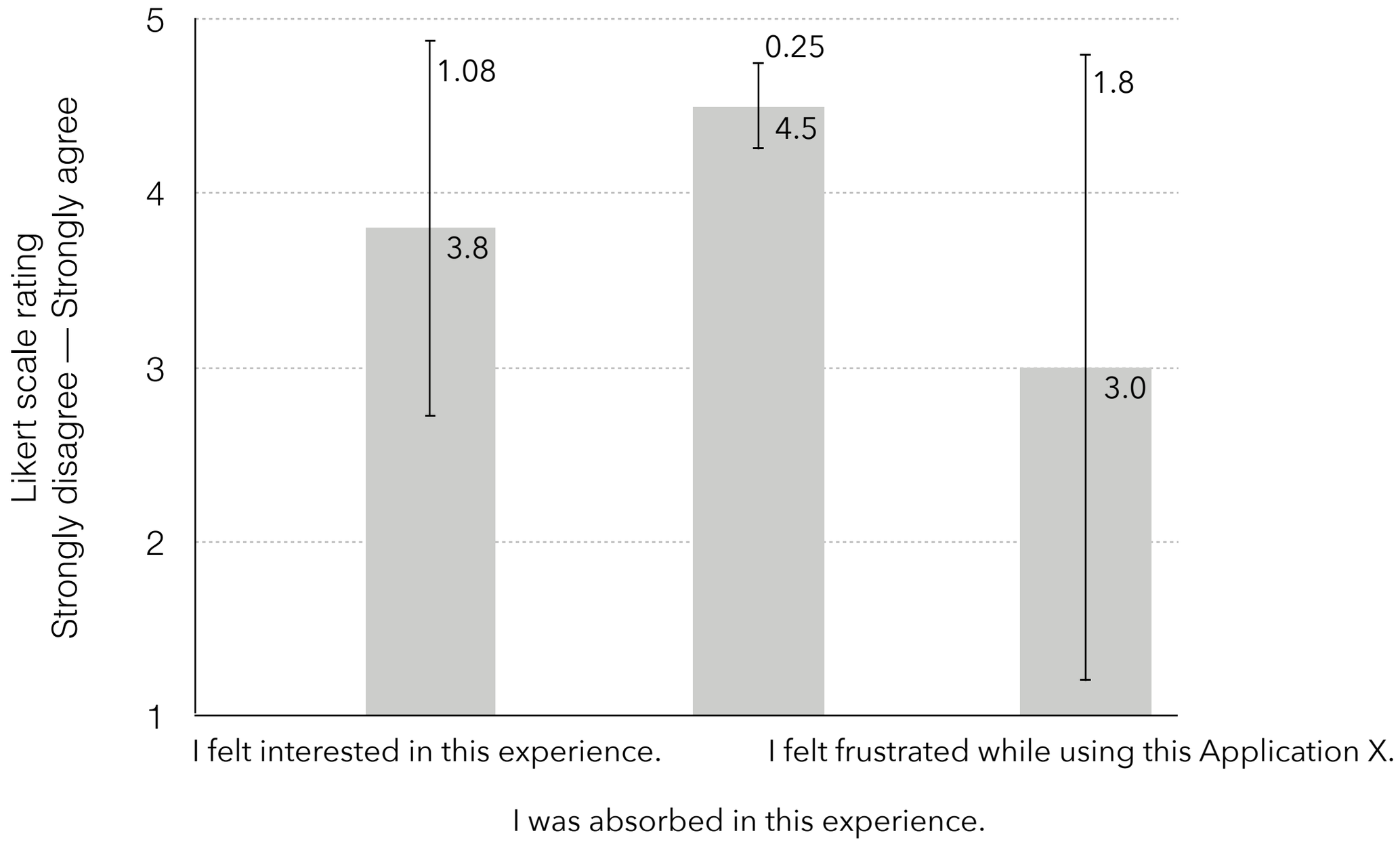
Mean and SD (calculation example)

```
var answers = [5, 3, 4, 2, 4, 4, 5, 2, 5, 4]; // n = 10


// average
var mean = (5 + 3 + 4 + 2 + 4 + 4 + 5 + 2 + 5 + 4) / 10; // = 3.8
```

```
// single squared deviations from mean
var d1 = (5 - 3.8)^2; // = ( 1.2)^2 = 1.44
var d2 = (3 - 3.8)^2; // = (-0.8)^2 = 0.64
// ... to d10



// mean of all single squared deviations
var variance = (1.44 + 0.64 + 0.04 + 3.24 + 0.04 +
                0.04 + 1.44 + 3.24 + 1.44 + 0.05) / 10 // = 1.16


// standard deviation
var standard_deviation = sqrt(1.16); // = 1.08
```

Mean and SD (calculation example)

Mean and SD (interpretation example)



Mean and SD (further reading)

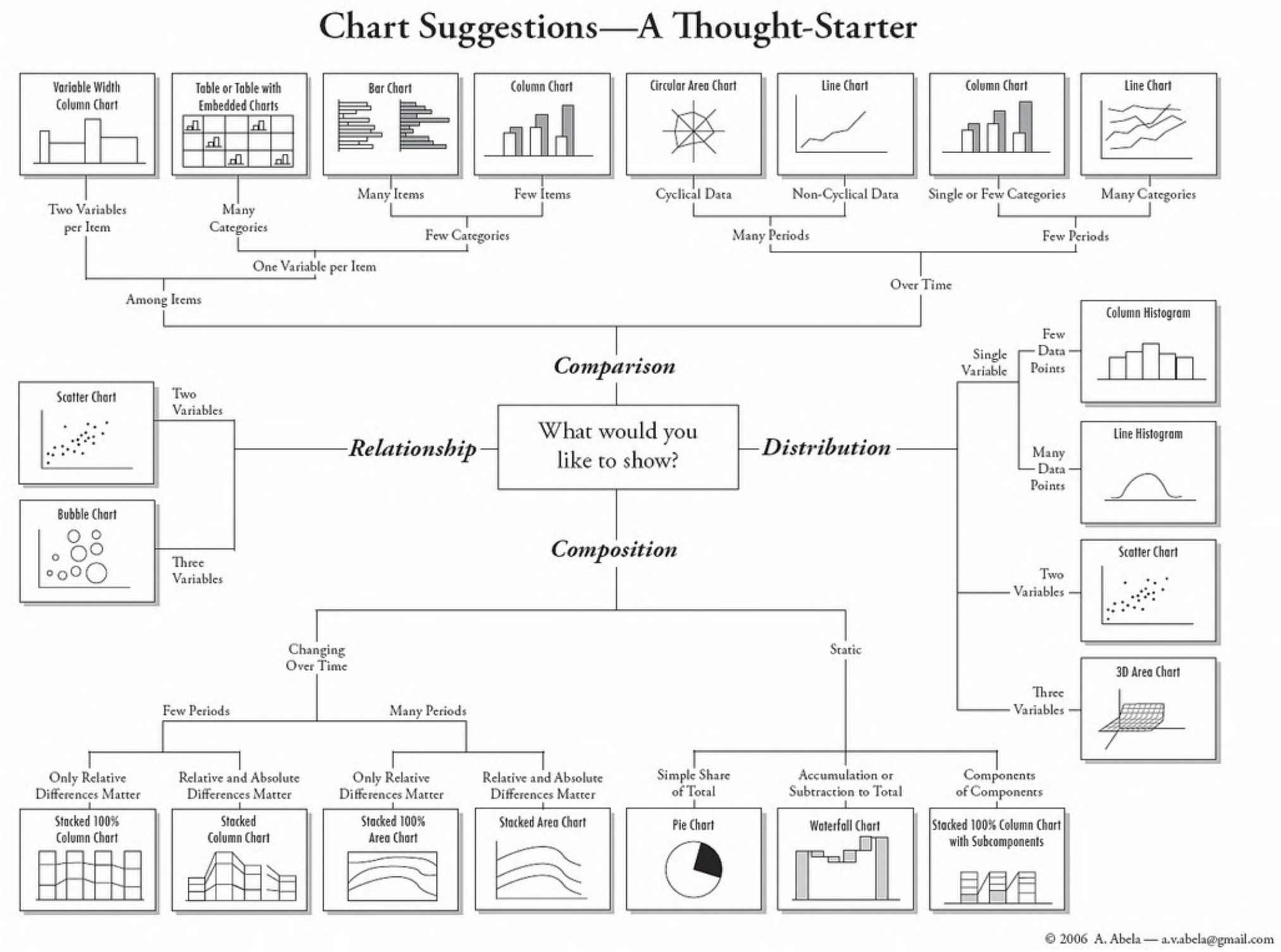
- Standard Deviations Formulas

  (further, detailed calculation examples, and difference between *population* and *sample*)

- **Note**: When using software to calculate standard deviation (e.g., R, MATHLAB, Spreadsheet, Excel, Numbers), make sure to use the correct standard deviation formula

  (*population vs. sample formulas*)

- Investigate other chart / plot visualizations, e.g., box plots (link 1, link2)



via FlowingData.com

System Usability Scale (SUS)

- “quick and dirty”, but reliable, tool for measuring usability
- 10-item questionnaire with 5 response options (Likert scale)
- reliable results on small sample sizes
- easily interpret the calculated scores of 0 - 100

  (the higher the number, the better it is)

via usability.gov

System Usability Scale (SUS)

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.

via usability.gov

System Usability Scale (SUS)

Adjective	Mean SUS Score
Worst Imaginable	12.5
Awful	20.3
Poor	35.7
OK	50.9
Good	71.4
Excellent	85.5
Best Imaginable	90.9

via A. Bangor, P. Kortum, and J. Miller (2009)

System Usability Scale (SUS)

Table 8.6 Curved Grading Scale Interpretation of SUS Scores

SUS Score Range	Grade	Percentile Range
84.1–100	A+	96–100
80.8–84	A	90–95
78.9–80.7	A–	85–89
77.2–78.8	B+	80–84
74.1–77.1	B	70–79
72.6–74	B–	65–69
71.1–72.5	C+	60–64
65–71	C	41–59
62.7–64.9	C–	35–40
51.7–62.6	D	15–34
0–51.7	F	0–14

via J. Sauro and J. R. Lewis (2012)

User Engagement Scale (UES)

- quality of user experience: depth of an user's (cognitive, temporal, affective, and behavioural) investment when interacting with a digital artefact / system
- engagement is more than user satisfaction
- arguably, the ability to engage and sustain engagement in digital environments can result in positive outcomes

via O'Brien et al. (2018)

User Engagement Scale (UES)

- UES: 31-item questionnaire with 5 response options (Likert scale), also referred to as UES-LF (long form), revised in 2018
- 4 dimensions: aesthetic appeal (AE), focused attention (FA), perceived usability (PU), and reward (RW)

  (Attention: The original UES featured 6 dimensions!)
- UES-SF: short form, capturing core concepts of the revised UES as a 12-item questionnaire with 5 response options (Likert scale)

via O'Brien et al. (2018)

User Engagement Scale (UES-SF)

FA-S.1	I lost myself in this experience.
FA-S.2	The time I spent using Application X just slipped away.
FA-S.3	I was absorbed in this experience.
PU-S.1	I felt frustrated while using this Application X.
PU-S.2	I found this Application X confusing to use.
PU-S.3	Using this Application X was taxing.
AE-S.1	This Application X was attractive.
AE-S.2	This Application X was aesthetically appealing.
AE-S.3	This Application X appealed to my senses.
RW-S.1	Using Application X was worthwhile.
RW-S.2	My experience was rewarding.
RW-S.3	I felt interested in this experience.

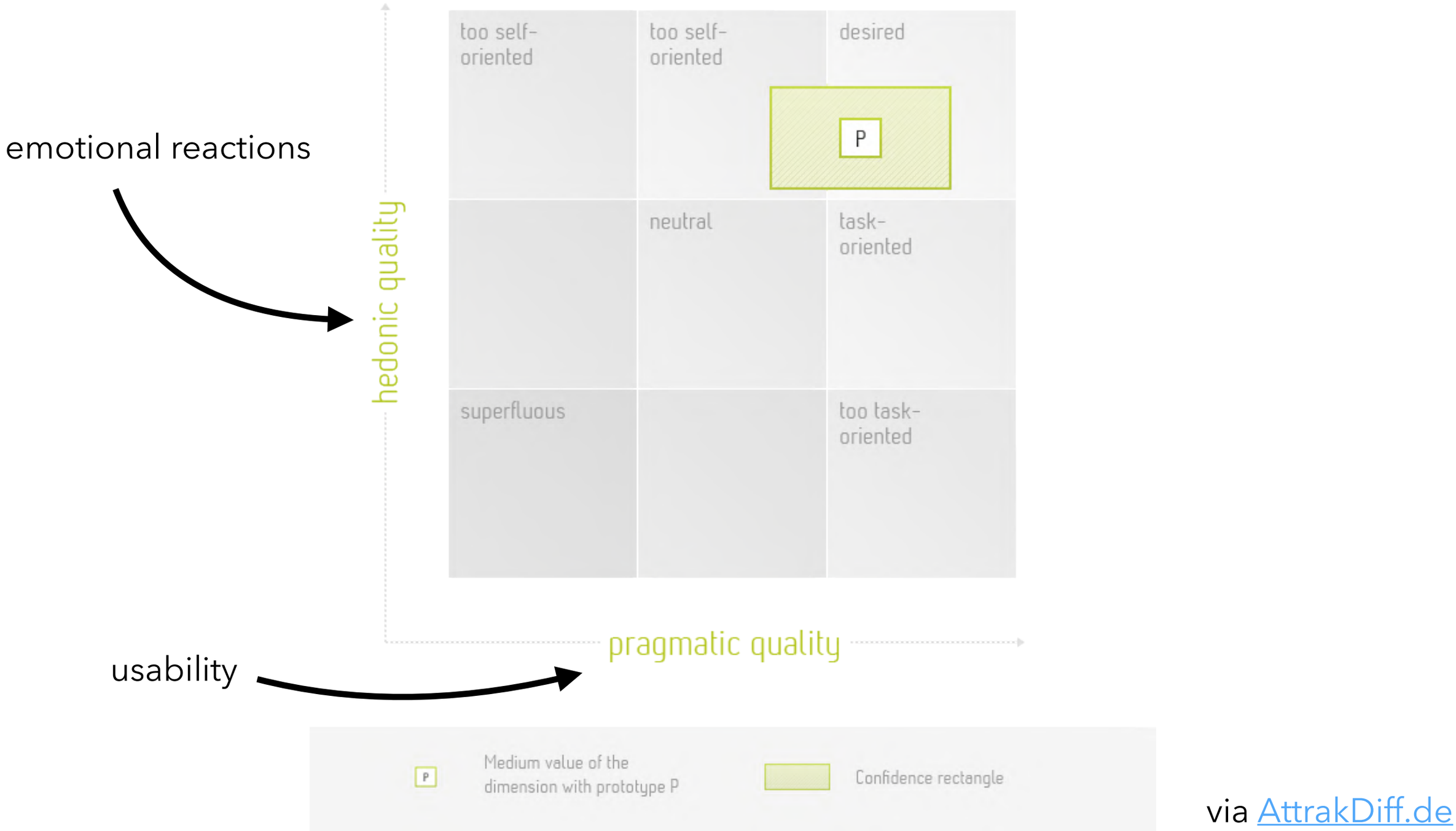
via O'Brien et al. (2018)

AttrakDiff questionnaire

- standardised approach to measure usability and design of a product
- online tool (registration required; free)
- different approaches possible
    - single evaluation
    - comparison A-B
    - Before-After

via AttrakDiff.de

AttrakDiff questionnaire



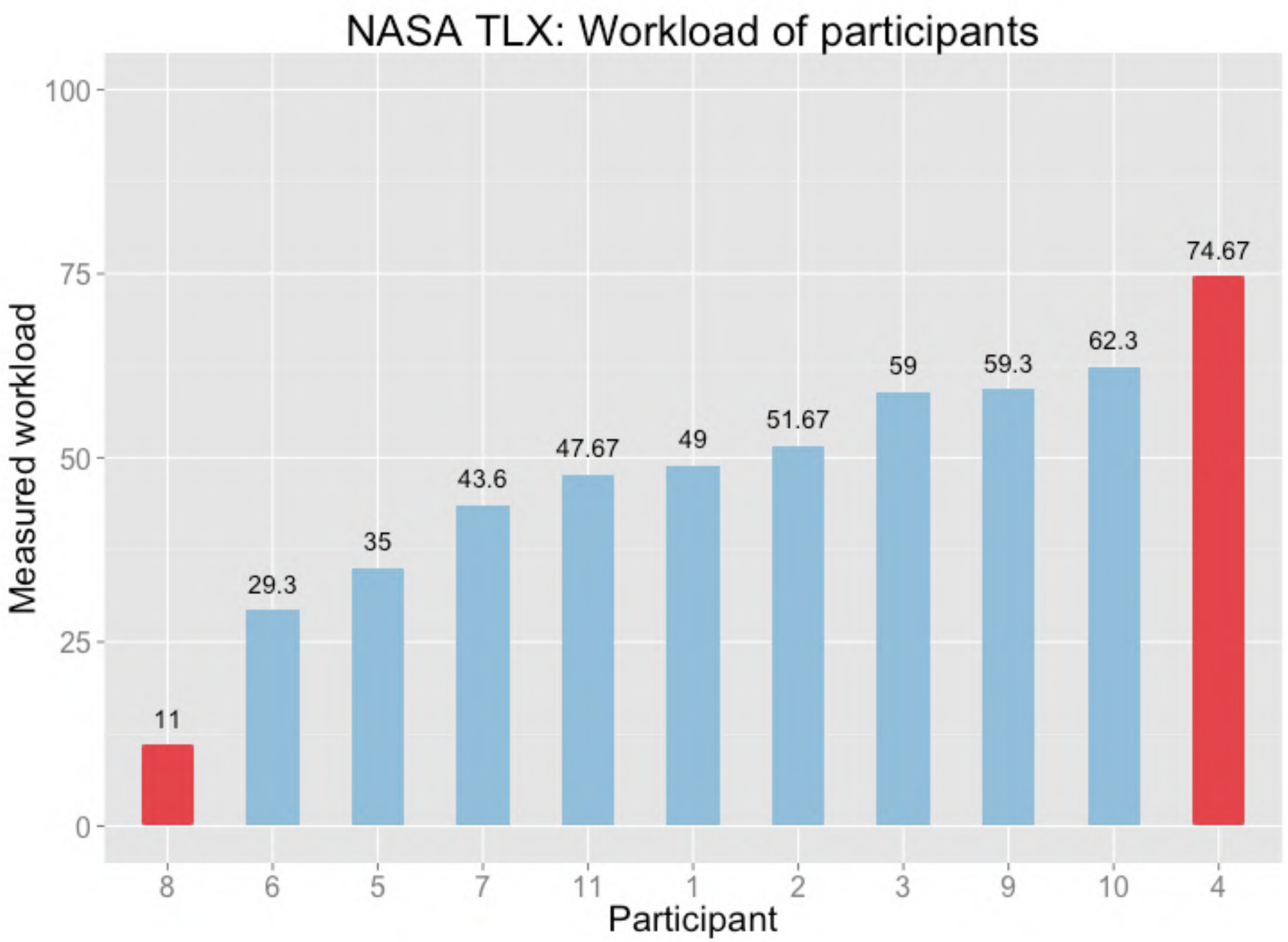
via AttrakDiff.de

NASA Task Load Index (TLX)

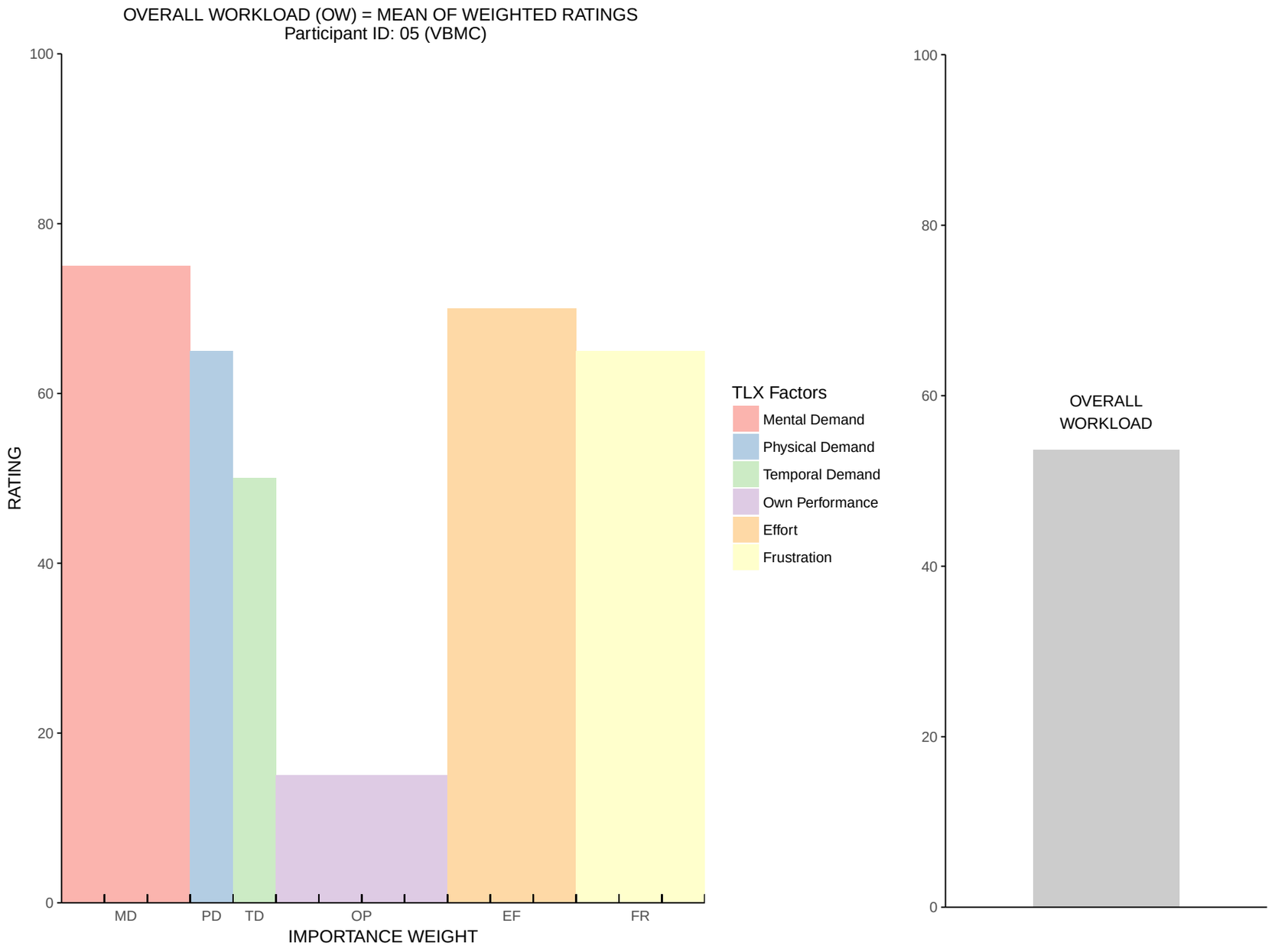
- 2-step approach, letting the participant first “weight” and then “rate” their interaction with a product
- 6 different factors, representing the **workload**
  mental, physical and temporal demand, their performance, effort, frustration
- analyse and estimate the interaction and interface design, providing indications if the participants felt e.g. bored, neutral or overburdened

via S. G. Hart and L. E. Staveland (1988), S. G. Hart (2006)

NASA Task Load Index (TLX)



NASA Task Load Index (TLX)



via github.com/nicoversity/tlx-vis-r

Simulator Sickness Questionnaire

- analyse and estimate “comfortability” of simulators
- origin in aviation, but also applied in related/similar conditions, such as Virtual Reality (VR)
- standardized, 16 items of investigation

  (e.g. fatigue, headache, nausea, vertigo)

  rated on a scale

  None - Slight - Moderate - Severe

via Kennedy, Lane, Berbaum, & Lilienthal (1993)

Flow Short Scale (FKS)

- investigate the overall “flow” interacting with an application, operating a system, completing a task, …
- origin in Csikszentmihaly’s (1988, 2014) flow theory
- standardized, 16 items of investigation

  (smooth and automatized process, ability to absorb, concern,

  fit of skill and requirements)

  rated on a scale

  Not at all - partly - very much

via Rheinberg, Vollmeyer, & Engeser (2003)

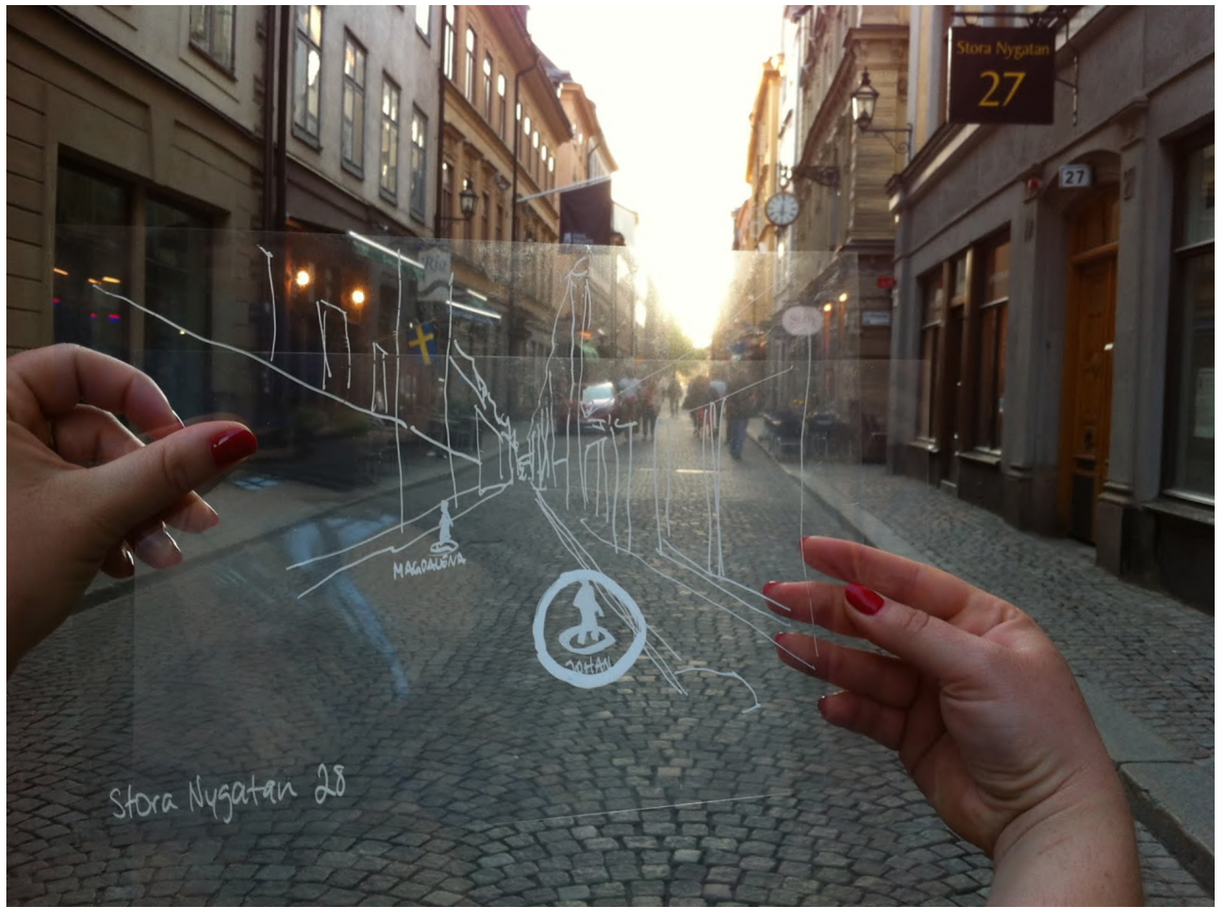
Explorative Expert Discussion

- dialog with experts knowledgable related to the context of your research objective in order to gather insights and opinions from their point-of-view
- present concept and idea, prototype (if already developed), in a semi-structured interview setting
- ideally with 2 to 4 experts at a time, enabling them to share different insights and start discussing views and opinions among themselves

Conceptual / Cognitive Walkthrough

- preparation of material, which is used to present and walk participants through the concept (of your idea)
- material can be sketches, paper prototypes, video, audio, presentations, …
- walkthrough should represent a “typical” session based on your idea (= user scenario)
- consider interaction and choices:
  structure your walkthrough in a way that the participant can decide between multiple options in certain situations
- ask questions / interviews = immediate feedback

Conceptual / Cognitive Walkthrough



History Explorer (4ME108-VT14)

Chase 'n' Race (4ME108-VT14)

Logging system

- “action-object-target” approach
  - each entry within the log file represents an event within the operation of the application
  - timestamp when the event occurred
  - the “action”, the “object” performing the action and potentially the “target”, the performed action is applied on

Timestamp	Action	Object	Target
1.000000	MOVE	Player	Stockholm
3.000000	TRIGGER	Filter_Menu	
7.000000	FILTER_APPLY	Connection_Area_Higher	Stockholm
11.000000	DISMISS	Filter_Menu	

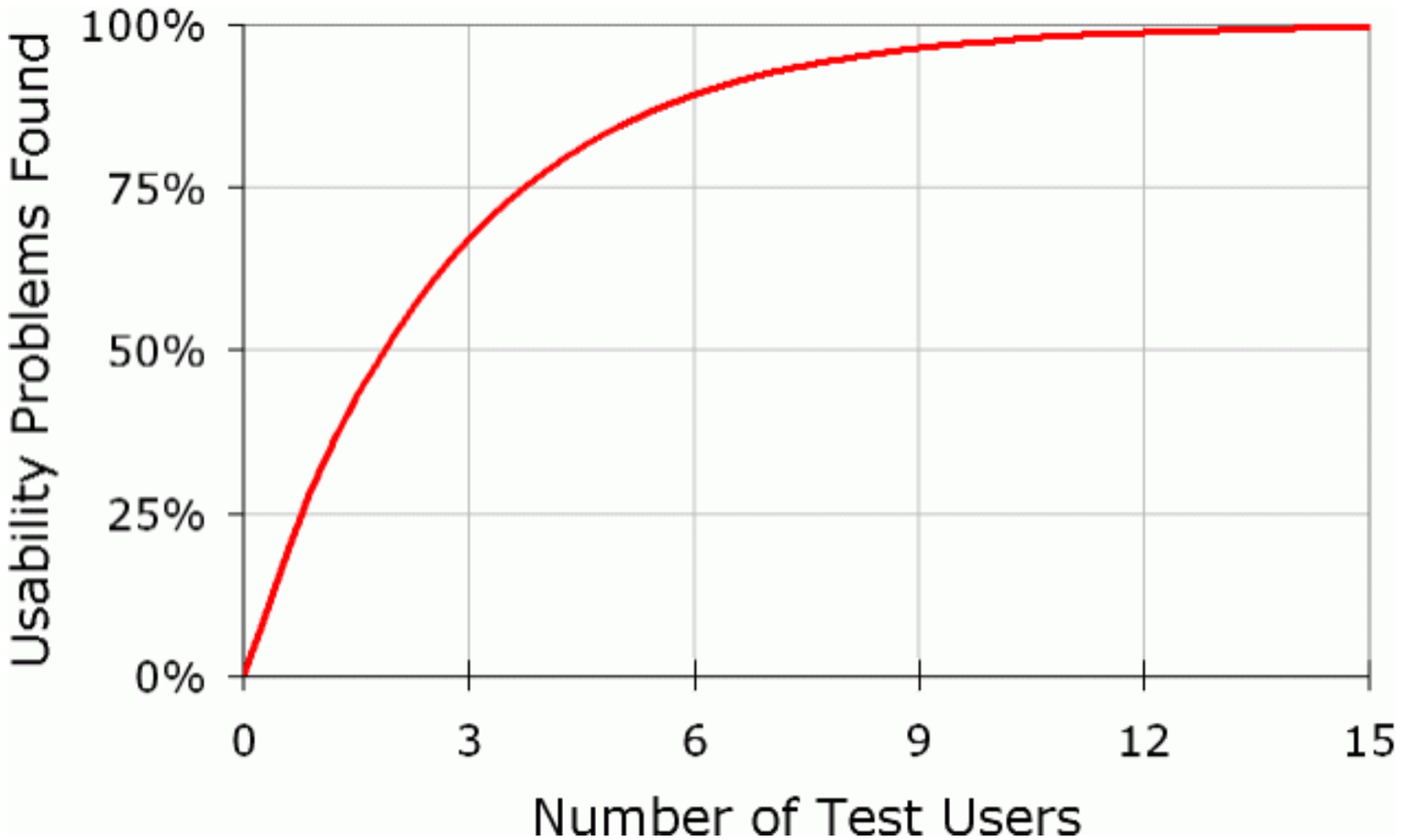
via github.com/nicoversity/unity_log2csv

Results / Analysis: Logging

Analysis			**Task 1**		**Task 2**	
			AVERAGE	**STDEVA**	**AVERAGE**	**STDEVA**
Average time spent in traveled City (in sec)			32.94	12.94	31.82	10.93
Amount of unique visited cities (max. 45)			11	4	11	4
Amount of visited cities			14	5	13	5
Amount of interactions			42	15	48	19
	Movement/Travels	**SUM**	15	8	15	8
		Successful	13	5	12	5
		Unsuccessful	2	3	1	2
		Forbidden	1	3	2	2
	Content Exploration	**SUM**	14	5	13	5
		Trigger	12	4	12	5
		Dismiss	11	4	12	5
		Rotation	2	3	0	1
	Filter Menu	**SUM**	13	8	20	11
		Trigger	4	2	5	2
		Dismiss	4	2	5	2
		Connection	8	6	13	10
		Area	0	1	8	6
		Population	6	4	1	2
		Reset	2	2	4	3
		Size	1	0	2	1
		Area	0	0	1	1
		Population	1	0	0	0
		Normal	0	0	1	1
Amount of time for completion (in sec)			421.84	160.61	391.74	142.33
in minutes			7.03	2.68	6.53	2.37

How many users do I need?

- 5 users, and run as many small tests as you can afford



via Nielsen Norman Group

How many users do I need?

- “5 users” - statement by Nielsen is controversial

  some researchers agree, some do not; in the context of usability testing

- overall, it highly depends on what and how you are going to test your product

  e.g. consider time, efforts, costs

  - Online questionnaire: ~ 30+ people
  - User interaction study on site: ~ 10+ people

Conducting a User Interaction Study

- 3 (5) phases
    - Preparation
    - Conduction
    - Analysis
    - Evaluation / Discussion
    - Conclusion

Preparation

- define goals of your user interaction study

  e.g. gain feedback about design and operation of a prototype

- identify the user target group

  e.g. teachers, students aged between 18 and 24 years

- construct (representative) tasks

- define data collection methods

  e.g. log files, SUS, interview

- schedule user interaction study and invite participants

Preparation

- **(technical) validation** of your prototype
    - make sure your developed prototype is operational

      e.g. no major bugs, user is able to complete a task, log files are working and recording
    - usually done with 1 - 2 participants

      who can then **NOT** be used again for your user interaction study
    - preparation of clear instructions / protocol of actions which the participant is asked to complete step-by-step

      (participant has no freedom of the actions!)

Conduction 1/3: Introduction

- welcome the user
- brief introduction to aims and purpose of the user interaction study
- explain formalities

  e.g. data is collected anonymously, consent to visually document (take pictures) the user interaction study, abort the study is possible at all times…
- answering potential questions of the user

Conduction 2/3: Execution

- provide and let the user complete tasks one by one
- conduct data collection

  observation and taking notes, audio/video recording, log files, document comments/questions of users…
- during the user interaction study, as a researcher you…
  - **should not** interact with the user
  - **should not** provide indications how the user performed with the task completion
  - **only help** if the user is really stuck and cannot continue alone

Conduction 3/3: Wrap-up

- post - user interaction study data collection

  e.g. user completes a self-constructed questionnaire, interview

- acknowledgement and sendoff

Analysis (of the collected data)

- bringing together notes/observations from all user sessions
- categorize identified problems and gathered feedback

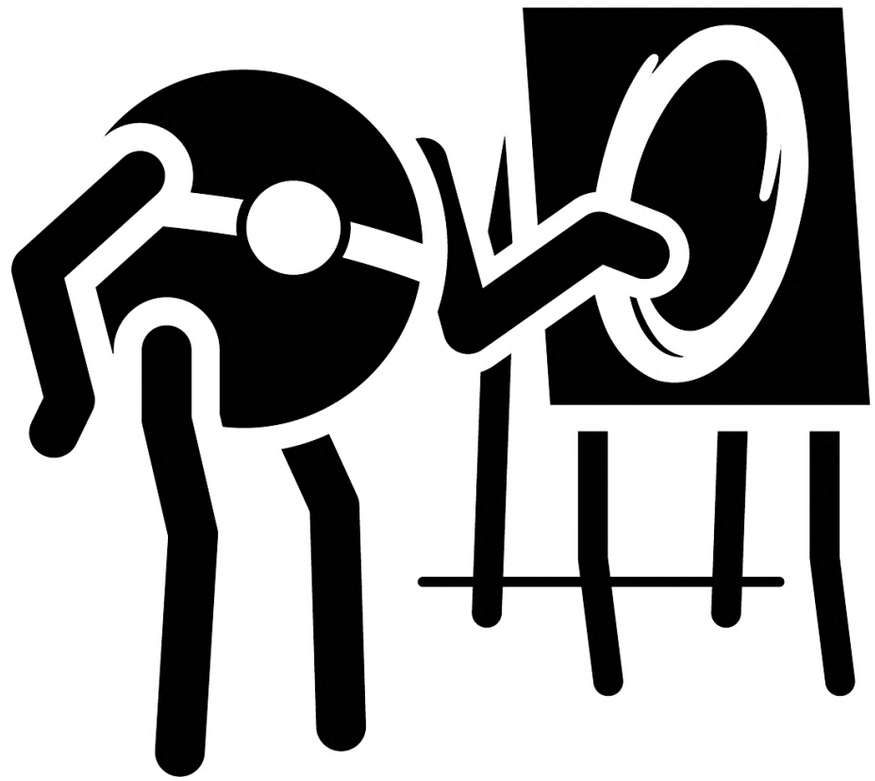
  e.g. layout and presented information, interaction, experience, hardware related, features
- analysis based on chosen tools / data collection methods

Analysis / Results vs. Evaluation / Discussion

- analysis / results = report facts, pure data,
  no meaning making !!!

- evaluation / discussion =
  putting facts into context, interpretation, meaning making,
  comparison to other studies / literature

Discussion: Your thesis ideas



References

- A. Bangor, P. Kortum, and J. Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, vol. 4, no. 3, pp. 114–123.
- V. R. Basili. 1992. Software modeling and measurement: the Goal/Question/Metric paradigm. Technical report, University of Maryland at College Park.
- H. L. O'Brien, P. Cairns, and M. Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, vol. 112, pp. 28-39.
- M. Csikszentmihalyi. 1988. The flow experience and its significance for human psychology. In Optimal experience: Psychological studies of flow in consciousness, M. Csikszentmihalyi (Eds.). Cambridge University Press, pp. 15-35.
- M. Csikszentmihalyi and Kanopy (Firm). 2014. Flow : psychology, creativity, & optimal experience with Mihaly Csikszentmihalyi.
- D. E. Grey. 2009. Doing Research in the Real World. 2nd ed, SAGE Publications Ltd.

References

- S. G. Hart and L. E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, vol. 52, pp. 139–183.
- S. G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908.
- R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal Of Aviation Psychology*, vol. 3 , no. 3., pp. 203-220.
- R. Likert. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, vol. 22, no. 140, pp. 1–55.
- J. Nielsen. 1994. Usability Engineering. Morgan Kaufmann.
- F. Rheinberg, R. Vollmeyer, and S. Engeser. 2003. Die Erfassung des Flow-Erlebens [The assessment of ow experience]. In Diagnostik von Selbstkonzept, Lernmotivation und Selbstregulation [Diagnosis of motivation and self-concept], J. Stiensmeier-Pelster and F. Rheinberg (Eds.). Hogrefe, Göttingen, pp. 261–279.
- J. Sauro and J. R. Lewis. 2012. Quantifying the User Experience. Elsevier.

Contact

Nico Reski

reski.nicoversity.com

@nicoversity

github.com/nicoversity

nico.reski@lnu.se

(PGP Key ID: B061D75B,
PGP Fingerprint: E826 C9FF 1701 0BAC
CA98 308C 6772 4499 B061 D75B)

Office: HUS D 2269 A

VRxAR Labs

Department of Computer Science
and Media Technology (CM)

Faculty of Technology

Linnaeus University, Växjö

Additional references

Portal icons in the presentation available via
bit.ly/portaliconpack