

HCI Design & Evaluation

Lecture 3 – User Confrontation (part I)

Monday 29-11-2021

Randy Klaassen

Timeline peer reviews:

Assignment	Hand in deadline	Receiving to be peer reviewed	Hand in peer reviews
Project Proposal	19 Nov	22 Nov	29 Nov
Concept video, persona, and scenario report	26 Nov	29 Nov	6 Dec
Interview report	3 Dec	6 Dec	13 Dec
Lofi prototype report	17 Dec	20 Dec	10 Jan
Hifi prototype report	14 Jan	17 Jan	24 Jan
Controlled experiment	16 Jan	17 Jan	24 Jan

**UNIVERSITY
OF TWENTE.**

Remember that formative feedback from the teaching team is leading.
Discuss the feedback from the peer-reviews with teaching team during the tutorials in case of doubts

Exam

- Friday 17-12-2021 8:45 – 10:15 (now: SC1)
- Multiple choice
 - Closed book, no other materials, no phones
- Example questions will follow soon
- Study material:
 - Lazar Chapter: 1, 5, 8, 11
 - Scenario Based Design paper
 - Value Sensitive Design paper
 - Preece Chapter: 2, 12
 - Klemmer: lecture 1, lecture 2 (video's)
 - Lectures

	Readings
(Fri 19-11)	Klemmer, Lecture 1. Available: YouTube Preece, Chapter 2 Link Lazar, Chapter 1 Link
persona and	Scenario-Based Design - Mary Beth LINK Value Sensitive Design and Information Systems - Batya Friedman LINK CHI - Guide to a Successful Video Submission
	Lazar, Chapter 5 Surveys Link Lazar, Chapter 8 Interviews Link Lazar, Chapter 11 Qualitative Data Link
	Preece, Chapter 12 Link Klemmer, Lecture 2. The Power of Prototyping Link Lazar, Chapter 15 Link Lazar, Chapter 16 Link
report	Lazar, Chapter 10 Link
17-12)	
	Lazar Chapter 2 Link Lazar, Chapter 3.1, 3.2, 3.3 Link Lazar, Chapter 4 Link Usability.gov Link ; Quant. Spec. Link
report	<i>provided during lecture</i>
experiment report	

The course

Week	Lecture (Location in Rooster)	on	Tutorial	Deadlines	Readings
1	Introduction HCI Course overview, topic and project	15.11	Ideation activity * Literature * Brainstorm * top 3 ideas	Proposal report (Fri 19-11)	Klemmer, Lecture 1. Available: YouTube Preece, Chapter 2 Link Lazar, Chapter 1 Link
2	Concepts, Ideation, Value Sensitive Design	22.11	Concept and Ideation * Scenario's, persona's, Values * Final #1 top idea * Video shotlist of above	Concept video, persona and scenario report (Fri 26-11)	Scenario-Based Design - Mary Beth LINK Value Sensitive Design and Information Systems - Batya Friedman LINK CHI - Guide to a Successful Video Submission
3	User confrontation	29.11	User confrontation * Interview script * Interview with peer groups * Interview results and conclusions	Interview report (Fri 03-12)	Lazar, Chapter 5 Surveys Link Lazar, Chapter 8 Interviews Link Lazar, Chapter 11 Qualitative Data Link
4	Lo-fi prototyping	06.12	Prototyping part I * (digital) prototyping * Setup user evaluation	-	Preece, Chapter 12 Link Klemmer, Lecture 2, The Power of Prototyping Link Lazar, Chapter 15 Link Lazar, Chapter 16 Link
5	Hi-Fi Prototyping	13.12	Prototyping part II * (digital) prototyping * User evaluation * Conclusions and results	Lo-Fi prototype report (Fri 17-12) Exam part I (Fri 17-12)	Lazar, Chapter 10 Link
	Experiment Design, Research methods and Data Analysis	20.12	Prototyping part III * Hi-fi prototyping	-	Lazar Chapter 2 Link Lazar, Chapter 3.1, 3.2, 3.3 Link Lazar, Chapter 4 Link Usability.gov Link ; Quant. Spec. Link
	Holiday				
	Ethics	10.01	Prototyping part III * Hi-fi prototyping * Experiment design + ethics	Hi-Fi Prototype report (Fri 14-01) Controlled Experiment report (Sun 16-01)	<i>provided during lecture</i>
	Guest lecture	14.01	User testing (Resits required)	Exam part II (Fri 21-01)	-
	n/a		n/a	Final project Report (Sun 23-01)	
	n/a		n/a	Oral Reflection (TBD) (option) resubmission Final project Report (Sun 06-02)	



Readings List:

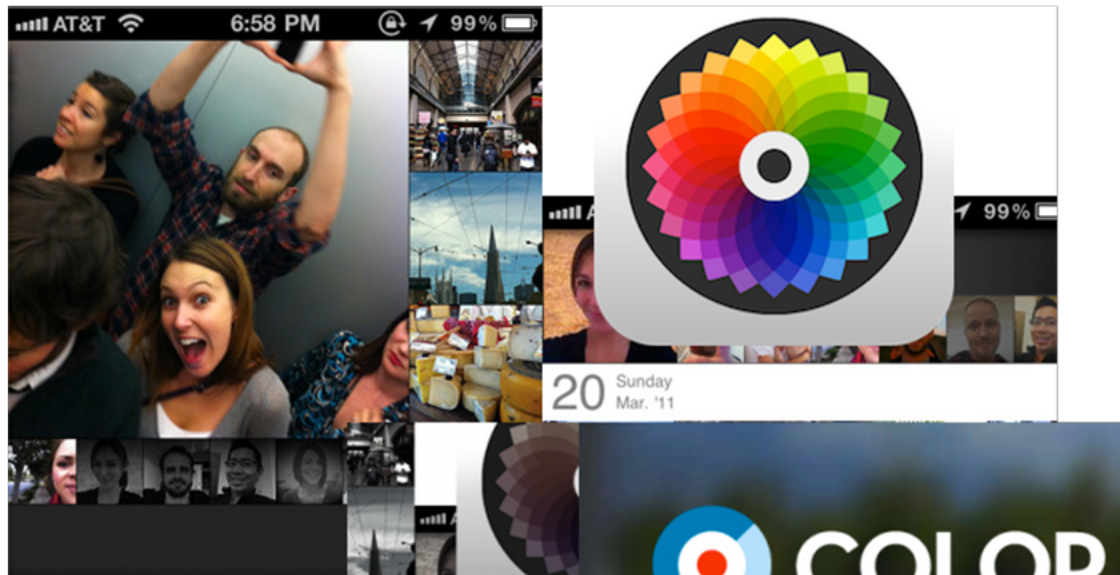
- Rosson, M. B. & Carroll, J. M. (2002) Scenario-Based Design. [LINK](#)
 Friedman, B., Kahn Jr., P. H. , & Borning, A. (2013) Value Sensitive Design and Information Systems [LINK](#)
 Preece, Sharp & Rogers (2019). *Interaction Design: beyond human-computer interaction*. 5th edition. John Wiley. [Link](#)
 Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann. [Link](#)

Today

- User confrontation (part I)
- Surveys (Ch. 5)
- Interviews (Ch. 8)
- Focus groups (Ch. 8)
- Analyzing results (Ch 8 & 11)
- This lecture will not cover everything, there is more in book!



Failed startups: color.com



- Over \$4 million in funding, able to secure domain and high profile ads
- Problem: no privacy controls for photos taken (all public)
- Did they do a survey or interviews? Maybe users would mention privacy...

Failed product: Google Glass

- Prominent foray into wearable technology
- Problem: \$1500 price tag, privacy concerns, public venues banned the device
- Did they speak with owners of public venues during development?
- Did they do user testing of prototypes in the wild and gather feedback from passersby?



most clients don't tell you what to make,
the user doesn't tell you what to make

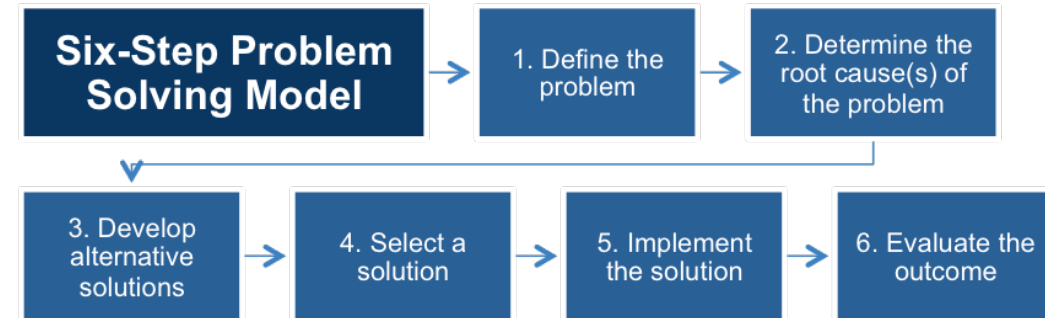
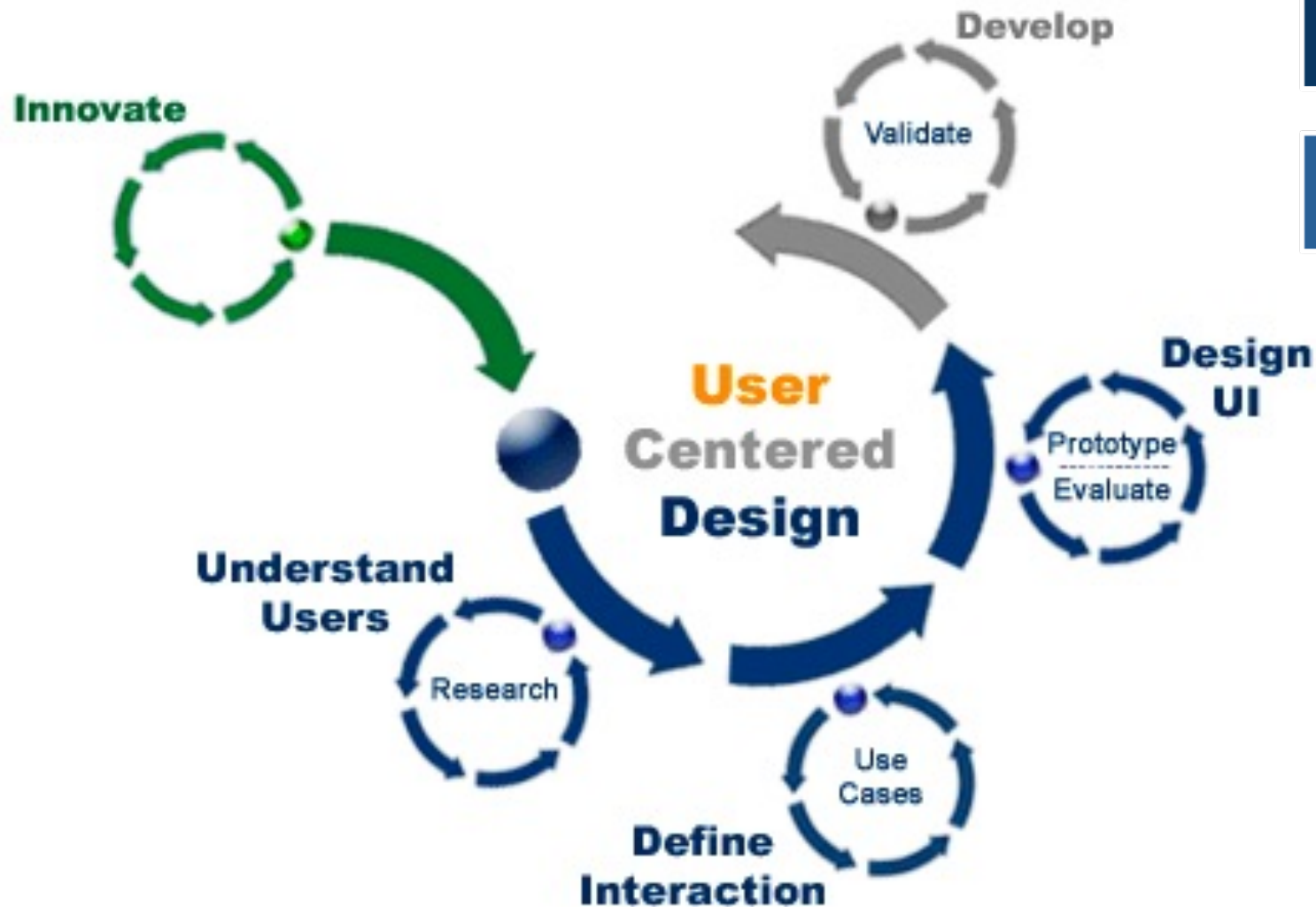
the users might not even know what they want
but neither do you



How?

- Survey
- Interviews
- Focus groups

Applications of Interviews and Surveys in HCI research



Learn by **DOING**.



Design Rationale:

*Grounded system design decisions are combined into a cohesive story about our product. This helps us to communicate clearly about **what we are making**, but also helps us reflect on whether we are **making the right thing**.*

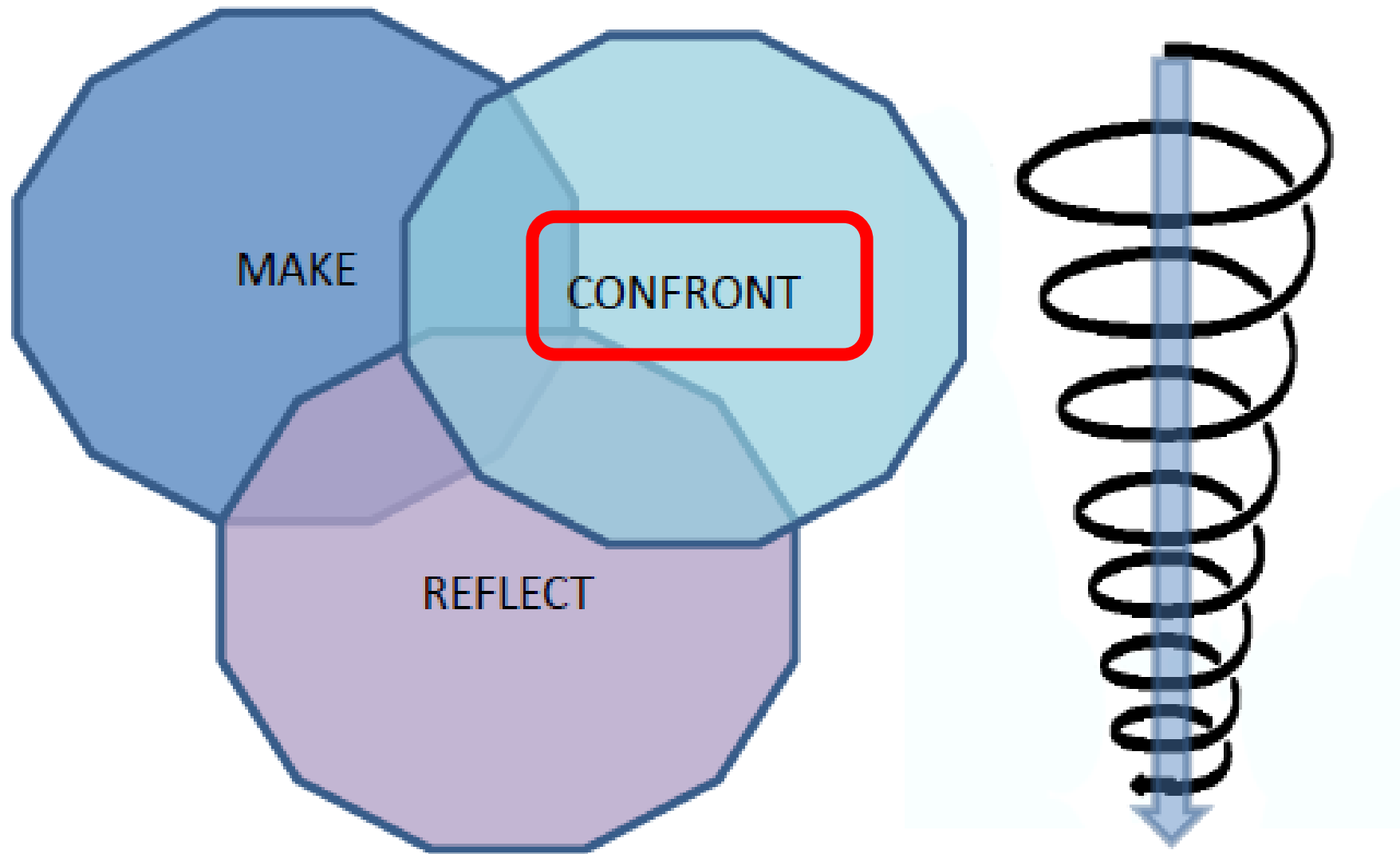


Figure 1: Activities and phases in the iterative method underlying our design work

Surveys (Ch. 5 Lazar)

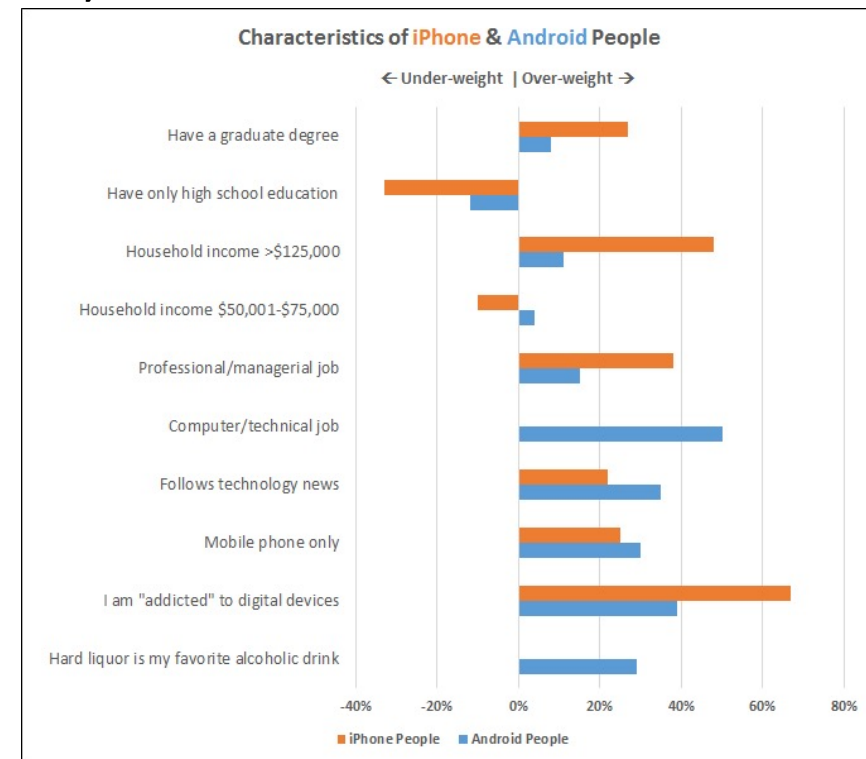
- Or questionnaires?
- Well-defined and well-written set of questions to which an individual is asked to respond
- Typically self-administered
- Strictly speaking:
 - Survey is the complete method
 - Questionnaires is the set of questions in this method

Survey

- One of the most commonly used research methods across all fields
 - To describe populations
 - To explain behaviors
 - To explore new ideas
- Ability to get a large number of responses in a short amount of time.
- Capture the "big picture", but not as deep or in-depth as other methods (such as interviews or focus groups → follow-up questions)
- Easy to distribute via email, websites or paper

Survey

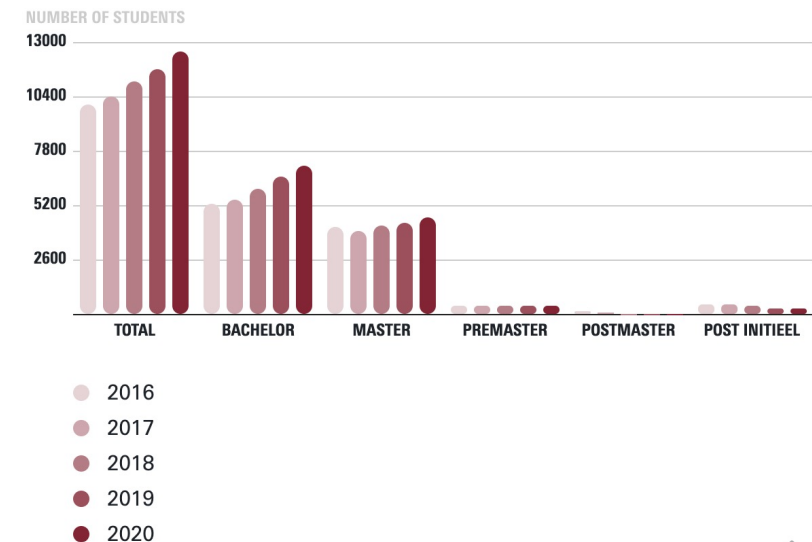
- Who are the targeted respondents of a survey?
 - Inclusion criteria
 - Targeted respondents (== user population, target users)
- People ≥ 30 years old?
- Software engineers?
- Smartphone users?
 - Android or iPhone?



Survey

- Where to find them and how to sample?
 - Phone books, email lists, social media, websites, ...
- Census (Dutch population) vs random/probability sampling (randomly select 500 students that were on Campus last week)
 - Random sampling of use (e.g. 500 UT students on Canvas)
- Stratification: stratified random sample, equal number of respondents invited to participate from each subpopulation.

STUDENT NUMBERS PERIOD 2016-2020



Survey

- In HCI we use more often nonprobabilistic recruitment
 - Opt-in (e.g. pop-up or link at website)
 - Snowballing (respondent recruits other potential participant)
- Without strict random sampling data is not valid? In the HCI community we consider it valid and acceptable.
- A HCI researcher typically collects the data themselves, from finding the users to analyzing the data.

Survey

- Errors:
 - Sampling error – not enough responses to make accurate estimations
 - oversampling
 - Coverage error – no equal chance of being selected (not on a list, not visiting a website)
 - Measurement error – poorly worded or biased questions
 - Nonresponse error – major difference(s) in demographics between people that responded and that were sampled (90% males responded and sampled was 50/50)

Survey Questions

- Open-ended questions:
 - Why did you stop using the Facebook app?
 - How did you feel about the usability of the Facebook app?
 - What barriers did you face when using the Facebook app?

- Close-ended questions:

What is your impression of using the website for www.veggieworld.com?

Please circle one number

Frustrating

Satisfying

1 2 3 4 5 6 7 8 9

FIGURE 5.1

A closed-ended question with an ordered response.

Source: QUIIS, see <http://www.lap.umd.edu/quis/>.

Which application do you use most often for text editing? (please select only one)

MS-Word

WordPerfect

Google Docs

OpenOffice Writer

WordPad

QuickOffice

Other (please specify)

FIGURE 5.2

A closed-ended question with an unordered response (single selection).

Survey Questions

What is your gender?

- woman
- man
- non-binary
- prefer not to disclose
- prefer to self-describe

(if the last option is checked a free form field opens up)

Figure 1. Our recommendation for how to ask about gender on large-sample HCI surveys.

Survey Questions

- Double-barreled questions
 - How long have you used the Word processing software and which features do you use?
- Negative wording
 - Do you agree that Word is not easy to use?
- Biased wording
 - Don't you agree that you should not use Word?
- Hot-button words
 - "liberal", "conservative", "racism", COVID-19, ...

Survey Questions - structure and order:

- Grouping of related questions
- Interesting questions at the start (to motivate completion)
- More boring questions or basic demographics at the end
- Sensitive or potentially objectionable near the end when respondent is already interested
- Contingent questions

Do you use, or have you used in the past, Microsoft Office 365 for e-mail?

Yes

No

If yes: Have you ever used the address book in Microsoft Office 365?

Yes

No

The diagram shows a survey question with two options: 'Yes' and 'No'. An arrow points from the 'Yes' option to a smaller box containing a follow-up question: 'If yes: Have you ever used the address book in Microsoft Office 365?' with its own 'Yes' and 'No' options.

FIGURE 5.4

A contingent question on a paper survey.

Survey – don't re-invent the wheel

Table 5.1 Survey Tools in HCI

Tool	Citations
Computer System Usability Questionnaire (CSUQ)	Lewis (1995)
Interface Consistency Testing Questionnaire (ICTQ)	Ozok and Salvendy (2001)
Perdue Usability Testing Questionnaire (PUTQ)	Lin et al. (1997)
Questionnaire for User Interaction Satisfaction (QUIS)	Chin et al. (1988) Slaughter et al. (1994) http://www.lap.umd.edu/quis/
Software Usability Measurement Inventory (SUMI)	http://sumi.uxp.ie/
Website Analysis and MeasureMent Inventory (WAMMI)	http://wammi.uxp.ie/

For more information about existing surveys for usability evaluation, the reader is encouraged to visit <http://garyperلمان.com/quest/>.

- And many more, e.g. SUS etc.
- Adapting is risky, tools are validated!
- Removing questions invalidates the tool
 - Adding questions at the end is the least bad ...

Survey – Data analysis

- Depends on probabilistic vs. non probabilistic survey
- How many responses
- Open-ended or closed-ended questions
- Quantitative vs qualitative data
- Data cleaning:
 - Double entries (ip, timestamps)
 - Incomplete (missing items)
 - Invalid (exclusion criteria)

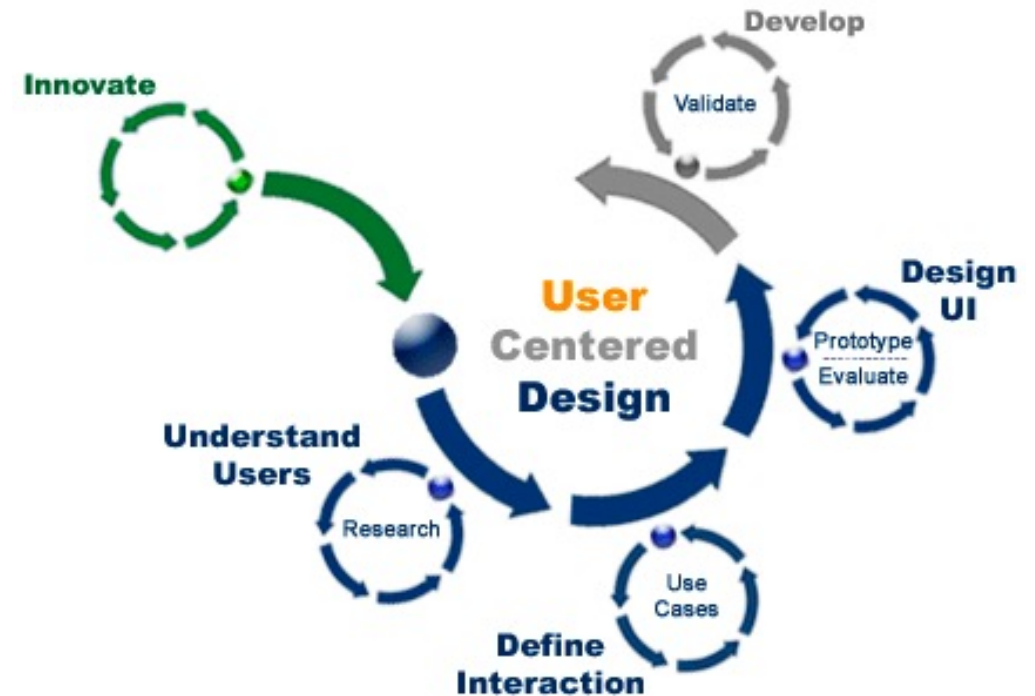
Interviews

- Direct feedback from interested individuals.
- What should a new tool do? → Ask the users.
- Does a proposed design do what it should do? If not, what should be changed or revised? → Ask the users.



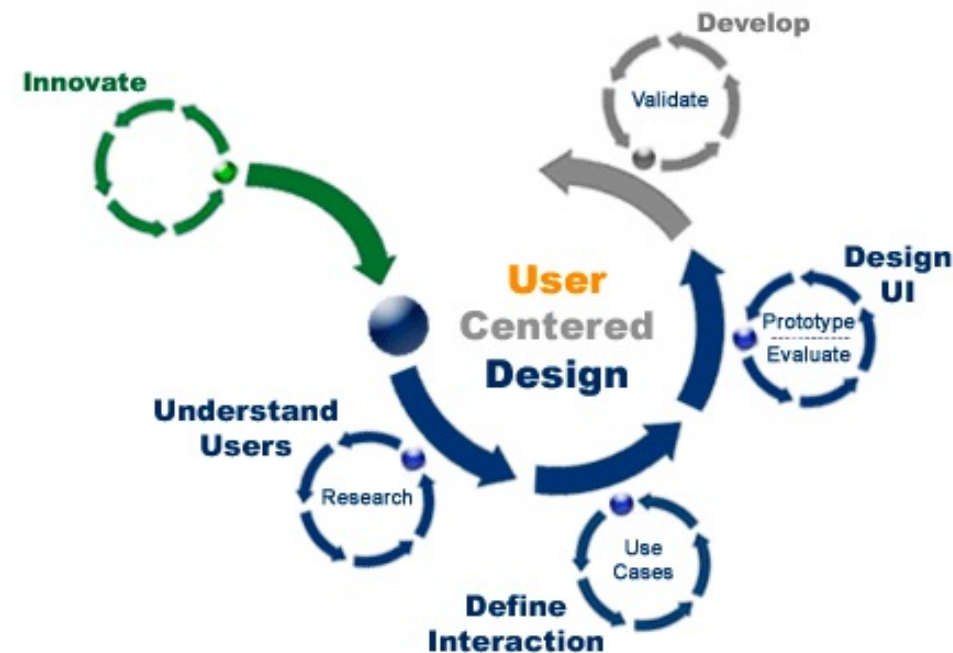
Interview stages

- Initial exploration
 - Understand user needs and goals
 - Don't focus on designs and solutions
- Requirement gathering
 - What are the goals?
 - How do they fulfil these goals?
 - What tools are there to this end?
 - Don't go for a narrow focus (it can blur the actual issues)
 - Sometimes low-level actions do play a role.



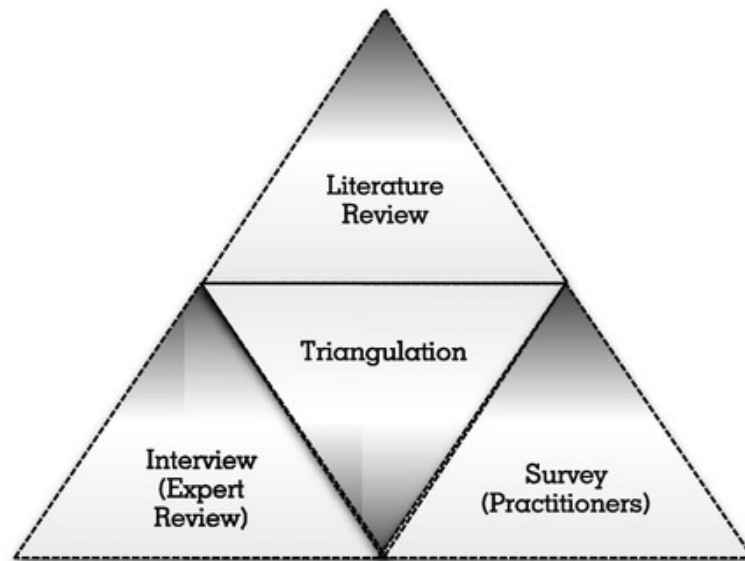
Evaluation & subjective reactions

- Complementary questions could help
- Results might contradict, e.g. the better version of your concept might not be chosen
 - dig-deeper, why, how and try to reconcile.



Interviewing

- Triangulation (combine with other methods)
- Example book:
navigation in-cars, first observations and then interview

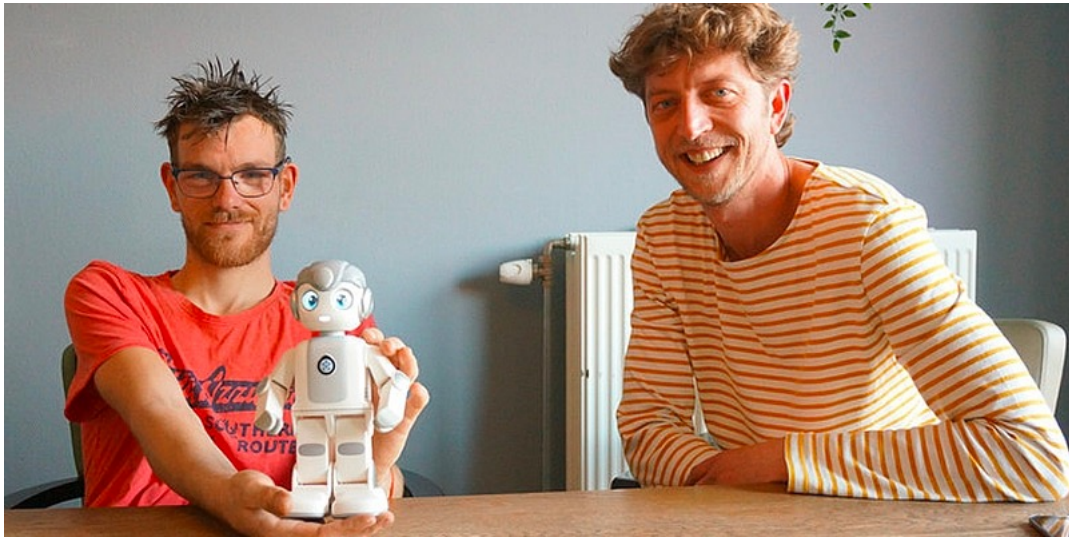


Interview strategies

- Fully structured
 - Relatively easy to analyze, strict script to follow, no room for extra questions and exploration
- Semi structured
 - Room to ask for clarification, possibility for exploration, harder to analyze, more skills needed
- Unstructured
 - Can go all over the place, more challenges in interpretation, more freedom

Types of questions

- Closed questions
 - limit users to a small number of predefined answers – yes/no, mc, T/F, Likert – easy to analyze
- Open questions
 - what do you think about X?
- Knowing how and for what you will use the answers
 - age (range, exact), education level, etc...
- Questions should be as simple as possible - don't use technical terms, jargon etc.



Awful



Not very good



Good



Really good



Brilliant



Interviewing

- [It is pretty intense and energy demanding to do]
- Take notes, also to come back to in the interview
- Consider various stakeholders (in the same domain)
- Properly executed interviews are time consuming,
 - you have to record/transcribe/write out the entire interview, verbatim-transcription (incl. every sound, entire durations)

Interviewing

Use tools while asking questions (blurring observation and interviewing)

- Probe— Technology probes are simple prototypes that demonstrate new ideas (Hutchinson et al., 2003)
- Key informants — individuals repeatedly called upon to provide important insights
- Contextual inquiry — in-depth interviews involving demonstrations of how participants complete key tasks (see Section 8.5.2)

What can go wrong with these questions?

“How were you feeling when you were using this software application?”

“recall bias”

What is wrong with this formulation?

“How long have you used the Word processing software and which advanced features have you used?”

→ “double-barreled question” asks two separate, and possibly related questions

“Do you agree that the e-mail software is not easy to use?”

→ Negative formulation (or worse double-negative)

What elements of the Persuasive Technology motivated you?

→ soft-version of “Hot-button” word

What can go wrong with these question?

“How often did you use the serious game?”



a persuasive game by Conceptlicios, ZGT & UT

What can go wrong with these question?

“How often did you use the serious game?”

<https://www.wooclap.com/XMKEMC>



a persuasive game by Conceptlicios, ZGT & UT

Customer Interviews

Good

vs.

Bad

FLIFFFT

Conducting an interview

- Planning and preparation
 - Pilot testing
- Interview script
- Recording the responses
 - Written notes – responses and own notes
 - Audio and/or video – what about privacy and GDPR?
 - What about performed tasks?
 - More roles?
 - Nonverbal behavior, backchannels



Conducting an interview

- Just use your common sense
- Be gentle and thankful for their time
 - Note “Subjects in the research may be outwardly hostile or simply indifferent to the project”, especially as a student.
- Debriefing

Conducting an interview

- Be gentle and thankful for their time
 - Note “Subjects of ethnographic research may be outwardly hostile or simply indifferent to the project” especially as a student.



Interviews vs Focus groups

- Interviewing is a labor-intensive data collection technique
 - 1 hour per participant, without data analysis!
- We can also meet with several participants in a focus group
 - 8 to 12 participants
- Limited doses of disagreement and debate can be very informative
- Focus groups are semi structured or unstructured

Interviews vs Focus groups

- Logistical and management challenges
 - Conversations take time, so less questions
- Participants can monopolize conversations
- Group dynamics



Ethnography (side note, not for exam)



Analyzing interview/focus group data

- Goal: to generate an accurate representation of interviewee response towards a holistic understanding
 - As soon as possible
- Possible biases in analysis of results e.g. cherry picking

Analyzing interview/focus group data

- Structured and closed questions vs. unstructured and open questions
- Notes, video, audio
- Verbatim transcriptions

- Content analysis – patterns, themes, other structure markers
 - Use quotes to support your themes

Qualitative vs. Quantitative data

- Qualitative Data
 - Observations
 - Quotes
 - Opinions
- Quantitative Data
 - Figures
 - Statistics
 - Log data

Research question(s) – reminder?

Research question(s) – reminder?

There are three general types of questions that we always find in research and design of User Experience.

- 1: A *knowledge question* is answered by collecting and analyzing information such as interviews, observations, or literature.
- Q: What are typical activities and experiences in a museum setting?
- A: The following activities and experiences: ...

Research question(s) – reminder?

2: An *evaluation question* is answered by doing an experiment or study with your product.

- Q: what is the technical performance of my product?
- A1: it can run uninterrupted without breaking for X hours consecutively.
- A2: It fails in the following conditions: ...

- Q: What is the impact of my solution on the daily practice of my users?
- A: It changes their experience in the following ways, ...
- Q: How does our solution change the user's social experience?
- A: ...

Research question(s) – reminder?

3: *A design question* is answered by making something.

- Q: What is a good algorithm for solving X?
- A: this one, because...

- Q: What is a good product for persuading people to practice their physiotherapy exercises?
- A1: the one I just made, because...
- A2: NOT the one I just made, because...

Research question(s) – reminder?

Sub-RQ's can also be related to **further specify your problem statement**. E.g:

- What aspect of [activity] do [user group] find particularly problematic when using [current solution]?

Sub-RQ's can be related to **specific** design questions you may have. E.g.:

- Which of these concepts (as displayed in your concept video) is more likely to motivate [user group] in [activity]?

Interviews

- RQ → variables you want to test
 - E.g. problem validation, design concepts
- Gather responses
 - Opinions
 - Observations
 - Ideas
 - Comments
 - What not to do
 - ??
- Don't ask RQs or sub RQ's directly to your users!

Analysing data - coding

- Emergent
 - Start by noting interesting concepts or ideas and refine while coding
- A priori
 - Uses a pre-defined structure to guide the coding of categories
- Not a simple choice
 - Probably you do both

Analysing data - coding

- Reliability issues in qualitative data analysis
- Ambiguous data:
 - The same word may have different meanings in different contexts.
 - Different terms or expressions may suggest the same meaning.
 - The data may be even more ambiguous when it comes to the interpretation of body language, facial expression, gestures, or art work.

Analysing data - coding

- Reliability checks span two dimensions:
 - Stability (intracoder reliability) – examine if the same coder rates the data in the same way throughout the coding process.
 - Reproducibility (intercoder reliability) - examine if different coders code the same data in a consistent way.
- the percentage of agreement among coders:

$$\% \text{agreement} = \frac{\text{the number of cases coded the same way by multiple coders}}{\text{the total number of cases}}$$

Analysing data - coding

- Cohen's Kappa: rates interrater reliability on a scale from 0 to 1. (with 0 meaning that all cases that are coded the same are completely by chance and 1 meaning perfect reliability)

$$K = \frac{P_a - P_c}{1 - P_c}$$

Table 11.4 Interpretation of Cohen's Kappa

Interpretation	Kappa range
Poor or slight agreement	$K \leq 0.20$
Fair agreement	$0.20 < K \leq 0.40$
Moderate agreement	$0.40 < K \leq 0.60$
Satisfactory agreement	$0.60 < K \leq 0.80$
Near-perfect agreement	$K > 0.80$

Analysing data - coding

- Subjective vs Objective coders (weird term)
 - Subjective or Inside coders those that developed, or made the coding scheme
 - Know how to label/code
 - Minimal training (they developed it)
 - - Constrained ability to think beyond (established concepts)
 - - Hidden meaning without being aware
 - Inflated reported reliability
 - Objective or Outside coders
 - More open to potential instances
 - (and see above, longer training, probably lower reliability)

Report what you do (subjective in HCI is okish)

Grounded theory

- Open coding
- Development of concepts
- Grouping of concepts into categories
- Formation of a theory

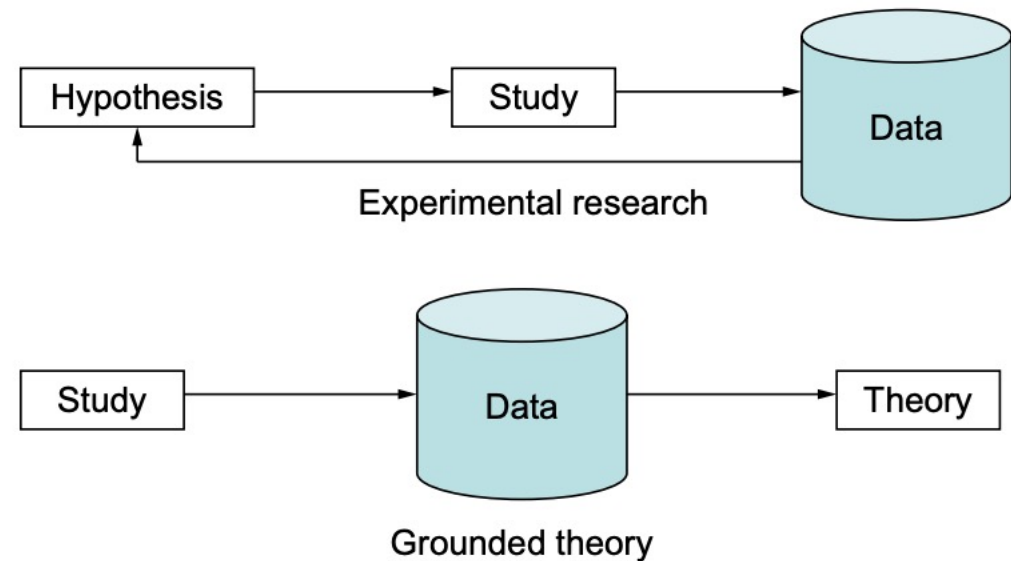


FIGURE 11.1

Experimental research compared with grounded theory.

Overview

1. Research questions (reminders on next slides)
 - Let them be checked by the teaching team before you go to step 2
2. Prepare your interview
 - Semi-structured questions
 - Define your roles during the interview
 - Let them be checked by the teaching team before you go to step 3 and 4
3. Find other groups to execute your interviews
4. Execute your interview (at least N participants, where N is the number of group members in your group)
 - Show your video at the end of the interview to avoid bias towards your problem statement!
5. Results and conclusions from your interview
 - How will it affect your concept?

For *your* interviews

- Use your notes and your recording
- Please don't write out the whole interview
- Use a type of coding: Emergent (possibly grounded) or A priori
- Write out quotes!

Presenting your data

- 4 out of 5 people were negative about X....
- So? Why? How?
- Use quotes!
 - Participant A mentioned that feature Y was frustrating because “...”

For *your* interviews

- Please adhere to the ethics/GDPR rules
- We have approval for this course from the committee if and only if
 - You do the interviews and evolutions during the tutorials with your peers
 - Your RQ's and interview (or later on activities, tasks, etc.) are reviewed by the teaching team before the activity
 - You ask oral informed consent from your participants
- More about this topic and the regulations at the UT in week 7

<https://www.utwente.nl/en/eemcs/research/ethics/>

https://www.utwente.nl/en/cyber-safety/privacy/faq_gdpr_index/

Thank you!