

Sonia Chiasson

COMP3008 – Fall 2020

Materials are intended for COMP3008 class use only.
Sharing or posting of class content is not permitted.

A photograph showing two people from the side, looking down at a white tablet device held by one of them. They appear to be in a modern office or home setting with a potted plant in the background. The person on the right is wearing glasses and a striped shirt.

Collecting Data and User Studies

Why do a user study?



The best way to learn about your design



Get feedback from users



Learn about things you hadn't considered



Understand users and what they want/need from your design

User studies

- Methods of data collection
- Types of user studies
- Methods and tips for running studies
- Analyzing and making sense of the data
- Reporting the data
- Making use of the data



Photo: 1141048370 by GaudiLab; shutterstock.com



Types of Data



Quantitative Data

Answers *How much?*

- Expressed as numbers (size, magnitude, amount)
- Measurable, statistics
- Can test specific characteristics or hypotheses

Qualitative Data

Answers *Why? What?*

- Detailed descriptions, subjective, “rich”
- Represents themes, patterns, stories
- Can be observed but not measured
- Open-ended enquiry



Photo: 521540095 by GaudiLab; shutterstock.com



Quantitative data

- Number of clicks
- Performance scores
- Task durations in seconds
- Likert scale responses
- Number of successful trials
- Number of computers you own

Qualitative data

- Area of expertise
- Explanation of why someone chose an item
- Description of a process
- Story describing someone's experiences with technology
- Someone's understanding of how a system works
- Favourite operating system

Subjective data

an individual's opinion, preference, experience, or judgment; not from some external measure.

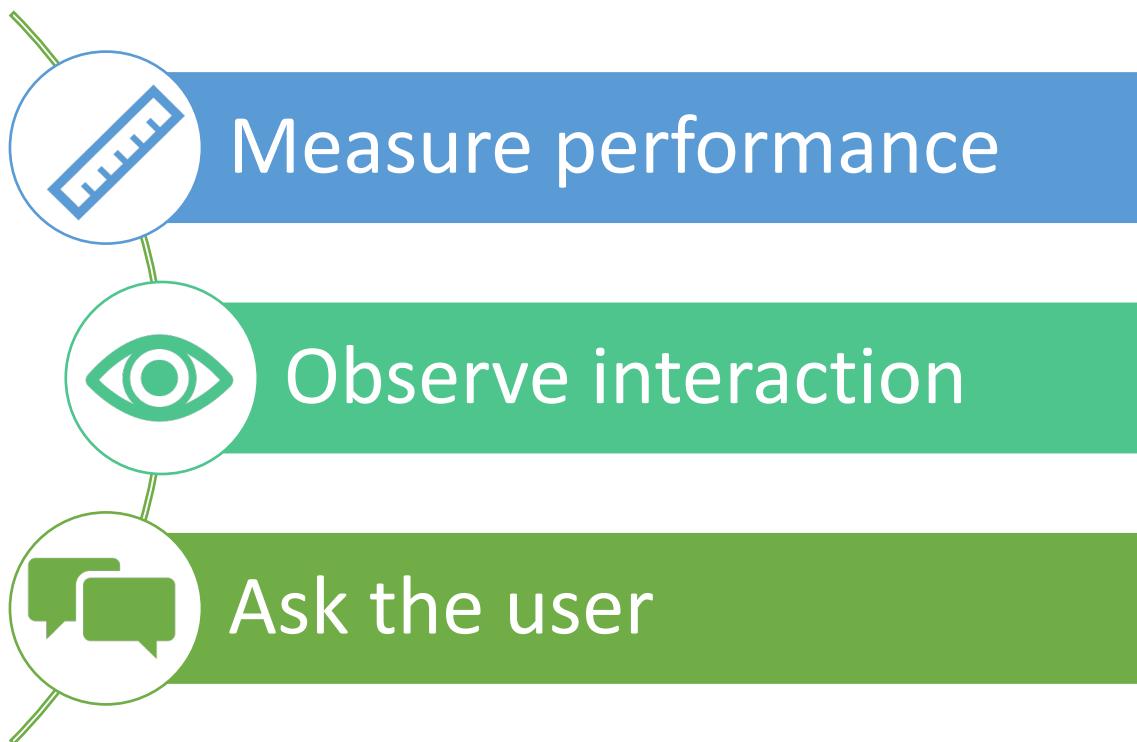
- Subjective-Quantitative: The user rates this system 8/10 for ease of use
- Subjective-Qualitative: The user thinks computers are too expensive

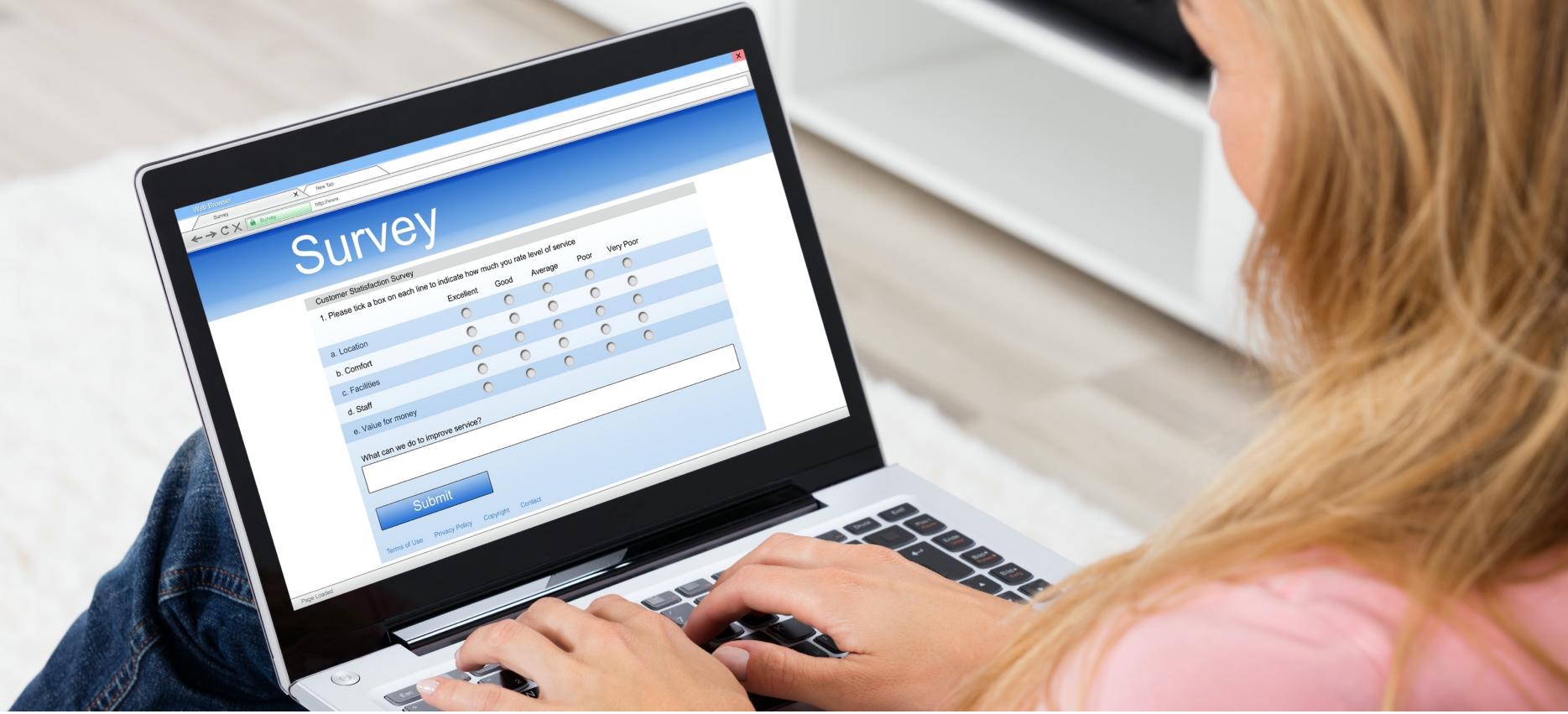
Objective data

'external to the mind' and concern facts and measurements.

- Objective-Quantitative: The user took 45 seconds to complete the task
- Objective-Qualitative: The user owns an Apple computer

Data Gathering





Questionnaires

Questionnaires



Collect user demographic data, opinions, and feedback



What people *say* is not always what they think, or do



Questions can be open or closed



Can be administered to large populations

Questionnaire design issues



Responses can be influenced
by order of questions



Provide clear instructions



Balance between
whitespace and
keeping compact

Graphic design
matters



Wording is very
important

E.g., decide if
phrases will be
all positive, all
negative, or
mixed

Multiple choice

Be specific when possible

- “1-3 times” instead of “rarely”

Cover whole range of possibilities,
without overlap

- 1-2 years, 3-5 years, 5+ years...
What if the answer is 2.5 years? Or
6 months? Or 5 years?)

Allows for categorical stats

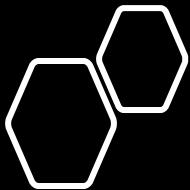
Which of these activities do you usually complete on the web? (check all that apply)

- banking
- income tax submission
- purchasing items

How often do you check your email?

- Several times per day
- Daily
- A few times a week
- Once a week
- Less than once a week

Likert Scales



Measures opinions, attitudes, beliefs

Asks about agreement

Typically use 3-10 choices

Odd = give option to be “neutral”,

Even = force users to “get off the fence”

Anchor at least the endpoints with labels (“Strongly agree”, “Strongly disagree”)

Easy to tabulate

Open to user interpretation (not equidistant)

I feel secure using my credit card online.

Strongly agree

Strongly disagree

1

2

3

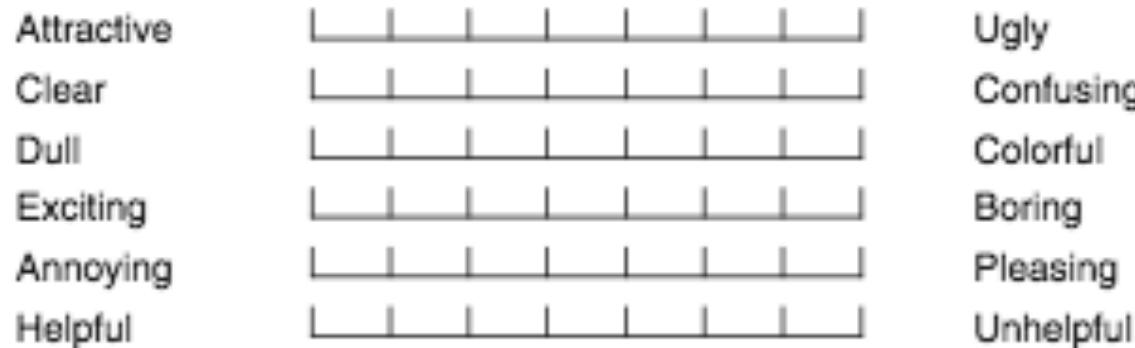
4

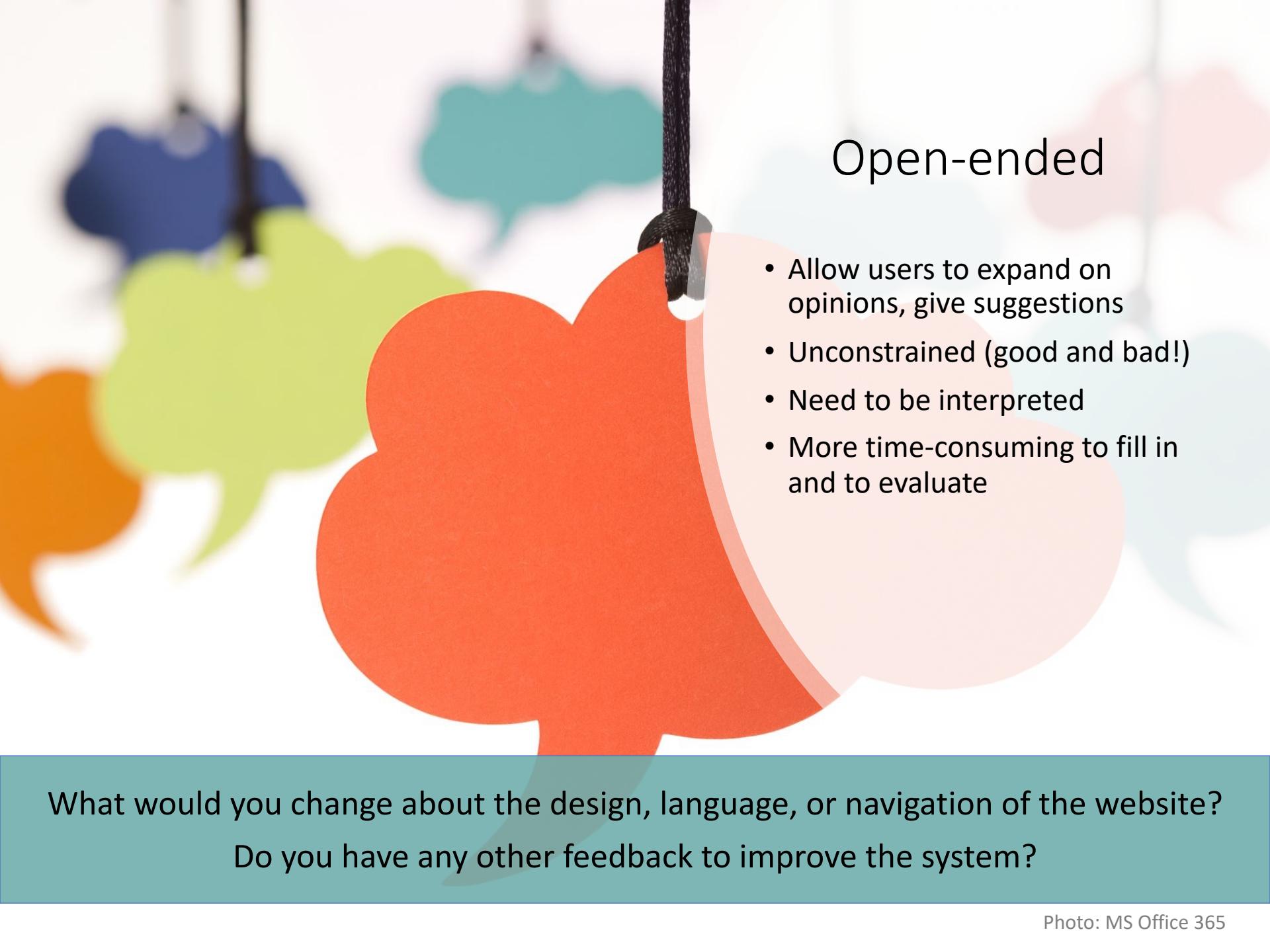
5

Semantic scales

- Anchored by two opposing poles
- Explore bipolar attitudes about a particular item
- Each pair of attitudes is represented as a pair of adjectives

CULearn is:

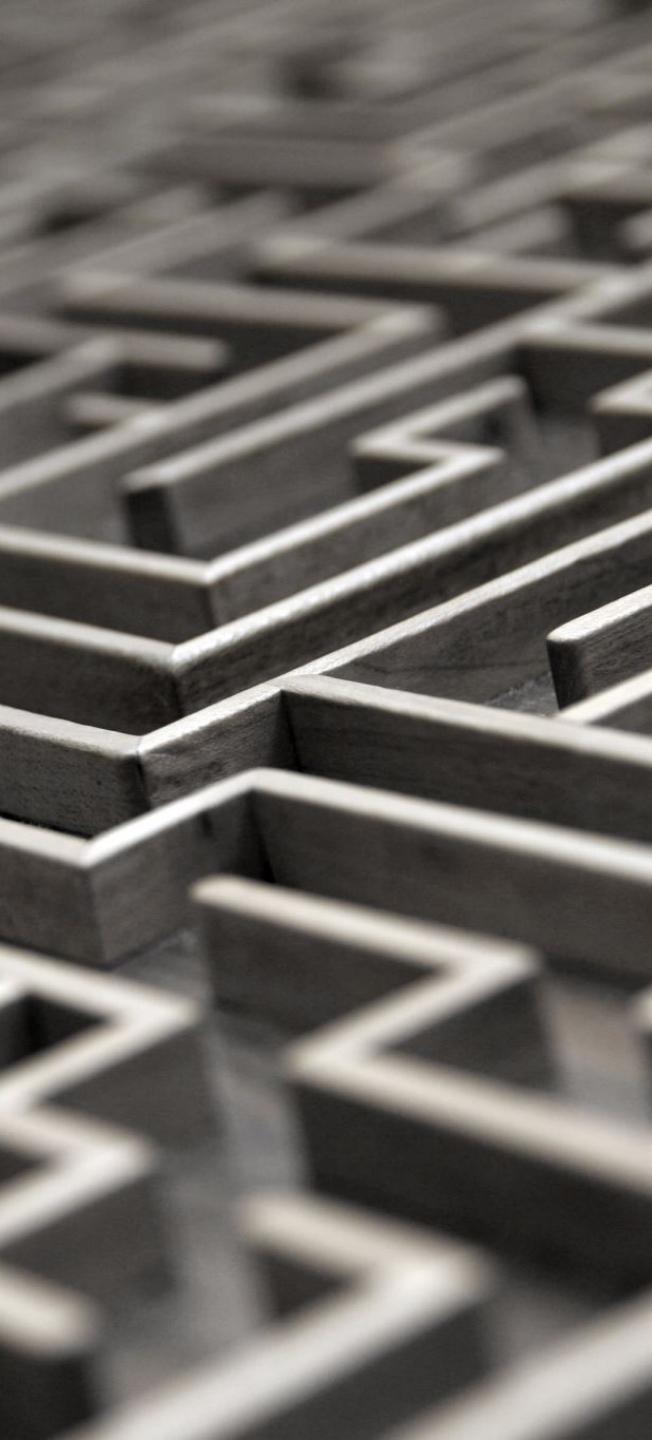


A decorative background element consisting of several large, colorful, paper-like speech bubbles in shades of orange, yellow, green, blue, and red, hanging from a dark string.

Open-ended

- Allow users to expand on opinions, give suggestions
- Unconstrained (good and bad!)
- Need to be interpreted
- More time-consuming to fill in and to evaluate

What would you change about the design, language, or navigation of the website?
Do you have any other feedback to improve the system?



General Advice

Be clear, concise, and avoid no jargon/slang

Avoid leading questions

Bad: Isn't this the best program you've ever used?

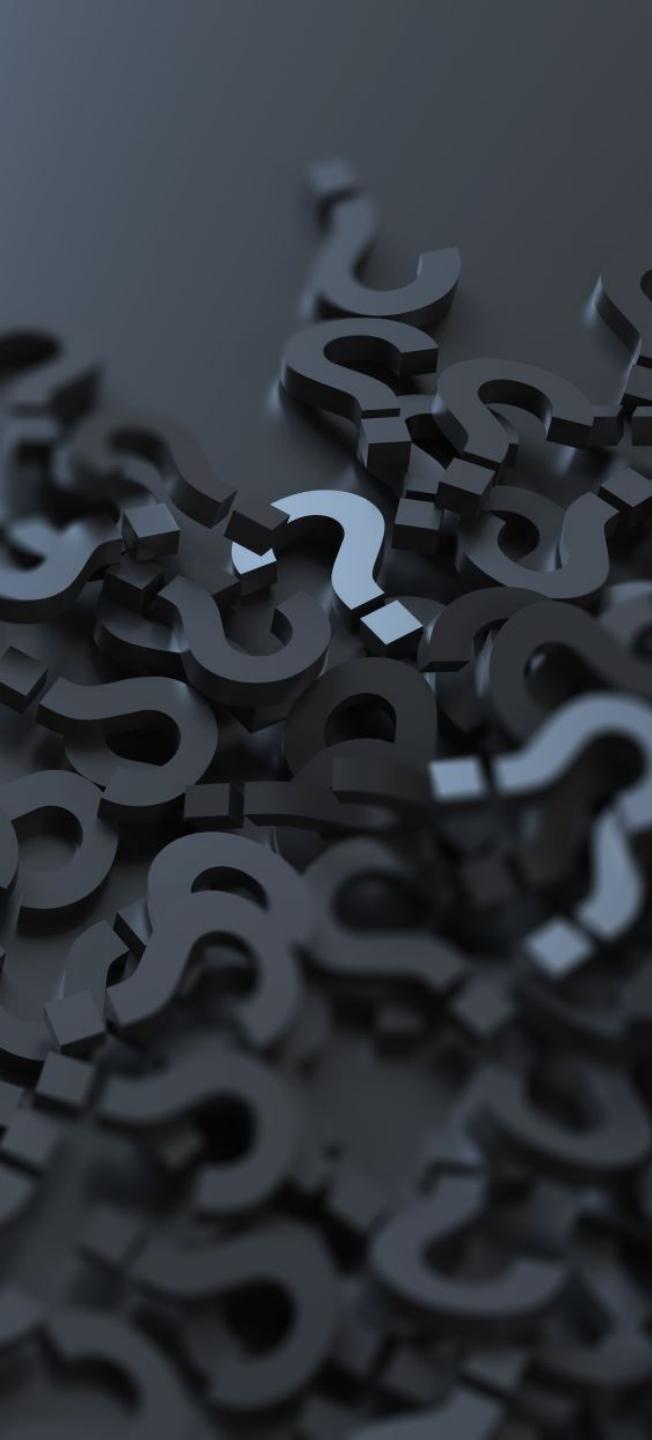
Bad: Is our new program better than the old one?

Bad: Some people say it is a waste of money to have computers available for all students at university. Do you agree or disagree?

Give equal weight to all multiple choice answers

Bad: Wonderful – Great -- Pretty good – Good – Bad

Good: Completely agree -- partially agree -- neither agree or disagree -- partially disagree -- completely disagree



General Advice

No double negative questions

Bad: Do you oppose not keeping the old system?

Bad: Do you agree or disagree that users should not be held responsible for security violations?

No double-barrel question

Bad: How satisfied are you with the layout and the wording of this page?

No “embarrassing” questions

Bad: Did you let your antivirus subscription expire?

General Advice



Keep it short



Use as secondary source

May not reflect “reality”, just what users perceive as reality or what they choose to reveal



Test your questions and validity

E.g., “How many passwords do you have?” ...
“Lots”, “I don’t know”, “10-15”, “5+”



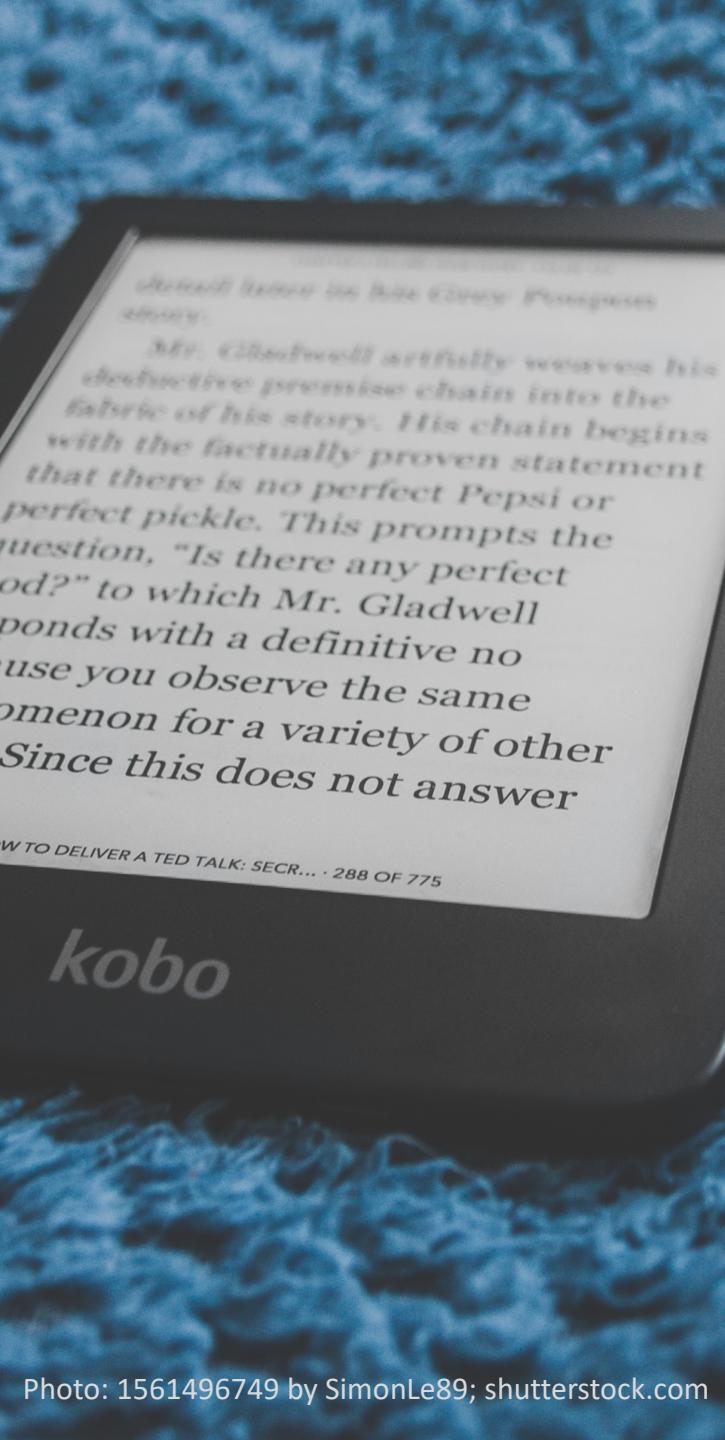
Is an existing validated questionnaire sufficient/available?

E.g., SUS, NASA-TLX, SEQ



Qualtrics, LimeSurvey,
SurveyMonkey, Google Forms

Consider privacy/security



Try it

You are user testing the Kobo e-reader website.

Improve the following questions.

1. Have you bought e-books, e-magazines, or electronic gift cards before?

 < 1 a month
 1-3 times a month
 3-5 times a month
 > 5 times a month
2. How much do you spend per year on e-books?
3. Is this the best e-book shopping interface you've ever used?
4. How often do you purchase or borrow e-books?

 < 1 a month
 1-3 times a month
 3-5 times a month
 > 5 times a month
5. Most people find the recommendations useful, Do you agree?



Observation



Pros and Cons

- + possible to do in a real-life context or in a lab
- + rich data
- + see what's happening for yourself

- observer's presence can disrupt work
- can be difficult to analyze or reproduce
- potentially time consuming



Direct Observation

User is given the task (or not), and evaluator just watches the user without interruption

Problem: no insight into the user's decision process, opinion...



Think-Aloud Method

Most widely used method in industry

User are asked to say what they are thinking/doing:

- what they believe is happening
- what they are trying to do
- why they took an action

Gives insight into what the user is thinking

Problems

- Awkward for the user (thinking aloud is not normal!)
- “Thinking” about it may alter the way people perform their task
- Hard to talk when concentrating on task

Co-Discovery Learning

Two people work together on a task

- Normal conversation between the two users is monitored
- Removes awkwardness of think-aloud, more natural
- Provides insights into thinking process of both users





Recording Observations

Paper and pencil (or laptop)

- cheap
- evaluator seems disengaged
- problem: writing is slow
 - prepared coding schemes can help; just tick off events

Audio recording

- capture discussion (think aloud, co-discovery)
- hard to synchronize with actions (e.g., interface actions)
- transcription is slow

Video recording:

- can see what a user is doing
(good to use one camera for user + screen capture)
- can be intrusive
- analysis can be challenging

Consider privacy



Try it

- Ask a friend to shop online for a new refrigerator while you observe.
- Ask them to think-aloud and write down your observations
 - Where did they start shopping?
 - What were their selection criteria?
 - Did they have any difficulties?
 - How did they decide which one to pick?
 - What else did you observe?

Interviews



“Conversations with a purpose”

Pros:

- excellent for pursuing specific issues
- address specific questions of interest
- more flexible than questionnaires:
probe more deeply on interesting issues as they arise

Cons

- accounts are subjective
- prone to rationalization
 - user’s reconstruction may be wrong
- time consuming to conduct and to analyze
- participants may have difficulty imagining ‘new way of doing things’
- evaluator can bias the interview



Types of questions

- Closed
 - “Do you own a smartphone?”
 - “How many times a week do you buy coffee at this store?”
- Open
 - How does buying coffee here compare to other places?
 - How would you improve the ordering process?
 - What frustrates you the most about ordering coffee?
 - What is most convenient about ordering coffee?
- Use simple language, no compound questions, remain neutral
- Order questions logically (but order may change)



Interviews

Structured

- Pre-determined set of questions

Unstructured

- More like a conversation, can often go into depth

Semi-structured

- Some pre-determined directions but flexible as situation evolves

In HCI, semi-structured and unstructured are most common

*Absolute key is to **listen** rather than talk: **practice silence!***

Sample Interview Flow

1. Introduce study purpose (carefully!), small talk, make person feel at ease, answer admin questions, set expected time
2. Get general participant background
3. (Letting off steam – let them bring up issues that they really want to tell you)
4. Cover remaining topics from your script
5. Finish with some easy, non-controversial questions
6. Wrap up, debrief

Interview Tips

- Make person feel at ease
 - They are *helping* you, you want their insight, there are no wrong answers, they are the expert
- Allow flexibility to deviate, answer out of order
- Practice first
- Avoid biasing participant, no leading questions, don't give your opinion
- Natural setting helps
- Participant does about 80% of the talking
- Prompt for more detail without leading
 - “what did you mean by that?” “can you tell me a bit more?”, “what does RTV mean?”
- Keep an eye on time – don't take longer than promised unless the participant says it's okay
- Usually audio-recorded



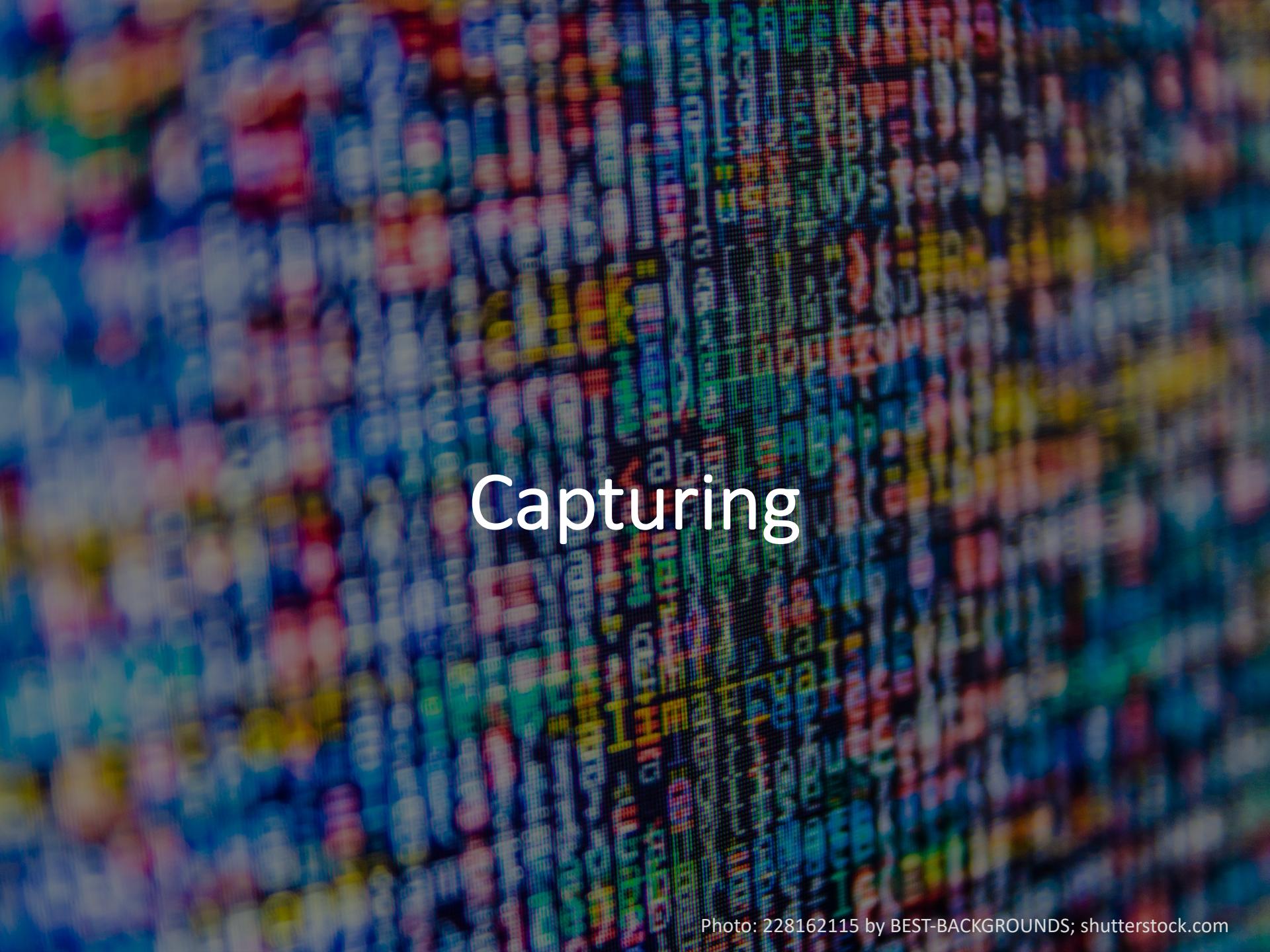
Focus groups (group interviews)

- 2 – 10 people interviewed at one time
- A skilled moderator is critical
- Usually recorded
 - + can accommodate **diverse and sensitive** issues
 - +/- opinions developed within a **social context**
some participants may be reluctant to take opposing view
 - + good way to identify “**proto-users**”: *most articulate, imaginative participants can help later with participatory design*
 - some interviewees may **dominate**

Try it

- Interview a friend or family member about their experiences with grocery shopping during the pandemic
- Use a semi-structured interview
 - Prepare a few questions but be ready to probe in other directions depending on what they say
 - Did they use technology?
 - What worked well?
 - What challenges arose?
 - What would have helped?
 - What concerns did they have?



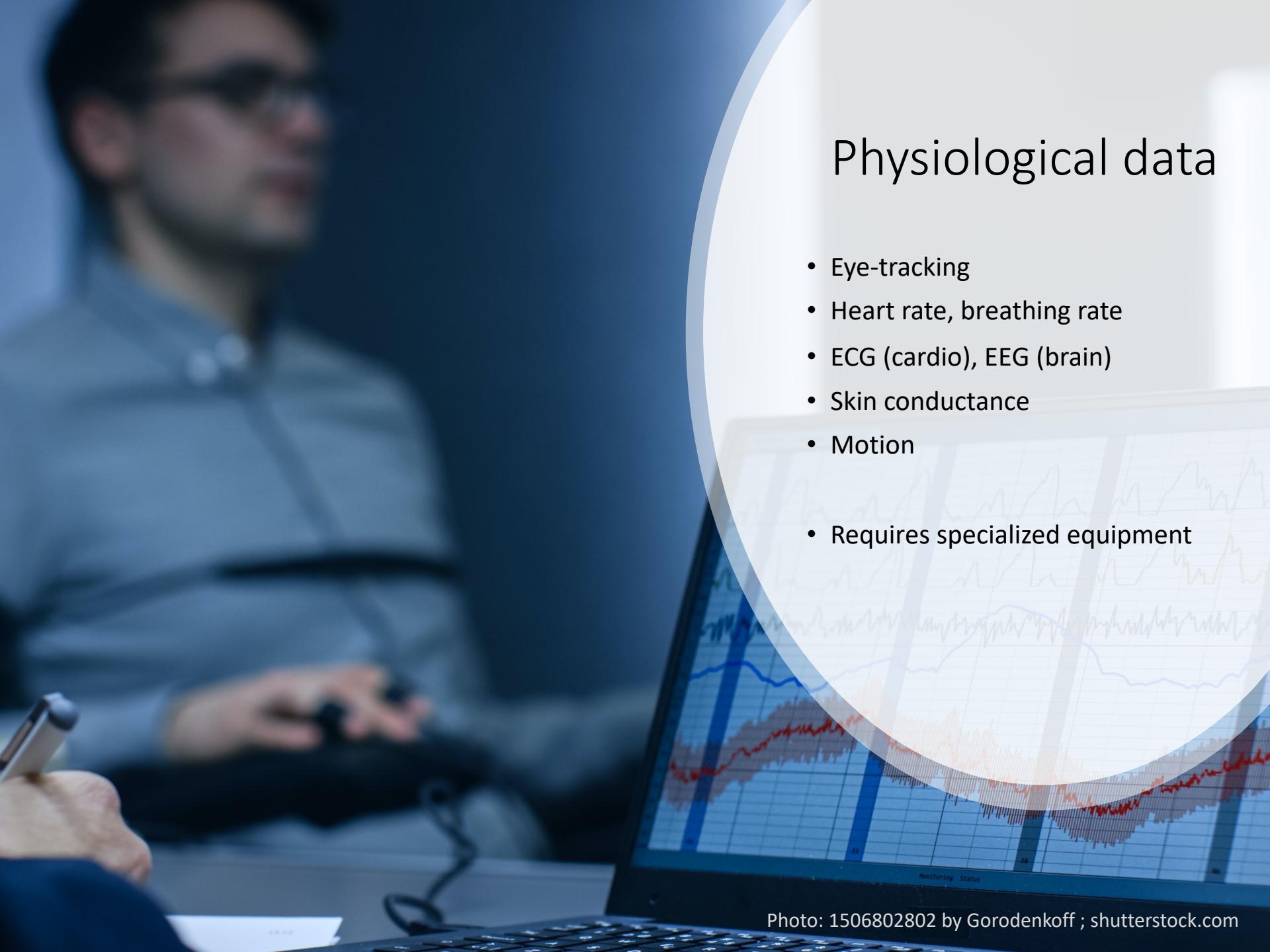
The background of the image is a heavily blurred digital display. It features a grid of small, colorful rectangles in shades of blue, green, yellow, and red, which appear to be individual pixels or data points. Interspersed among these are several larger, semi-transparent rectangular overlays containing faint, illegible text and graphical elements, possibly representing charts or specific data series.

Capturing



Performance data

- Timestamps, timing
 - User clicks, selections, data entries
 - Scores
 - Task outcomes
 - ...
-
- Software scripts, instrumentation, logging, metrics

A blurred background photograph shows a man in a dark shirt looking at a laptop screen. The laptop screen is visible in the foreground, displaying various physiological data graphs and waveforms in blue, red, and grey colors on a grid background.

Physiological data

- Eye-tracking
- Heart rate, breathing rate
- ECG (cardio), EEG (brain)
- Skin conductance
- Motion
- Requires specialized equipment

Hints for capturing data

- Test carefully ahead of time
 - Are you capturing everything you need?
 - Does the output make sense?
- Pay close attention to formatting
 - What will facilitate analysis?
- How are you managing different users, tasks, trials?
- Consider security/privacy

Try it

- Think through how you would collect data about prototype user authentication system you are evaluating.
 - What kind of data do you need to collect?
 - How should you record and store it so that it's easy to do your analysis later?
 - Are there any privacy or security concerns with the data you're collecting?





User Studies

A close-up photograph of two men looking down at a tablet device. The man on the right has dark hair and a beard, wearing a light blue button-down shirt. The man on the left has light brown hair and a beard, wearing a dark shirt. They are both focused on the screen of the tablet. The background shows wooden window blinds.

Lab-based / Controlled

Characteristics:

- Controlled environment
- Direct observation, “think-aloud”

Used for:

- Measuring performance & improvements
- Discovering usability problems
- Gaining confidence of design before riskier testing

Potential Drawbacks:

- Realism can be difficult to achieve
- *Novelty effects* (hey, this is cool!)



Field-based

Characteristics:

- Deployed in regular setting
- Maybe more ecologically valid
- More expensive and time-consuming

Used to:

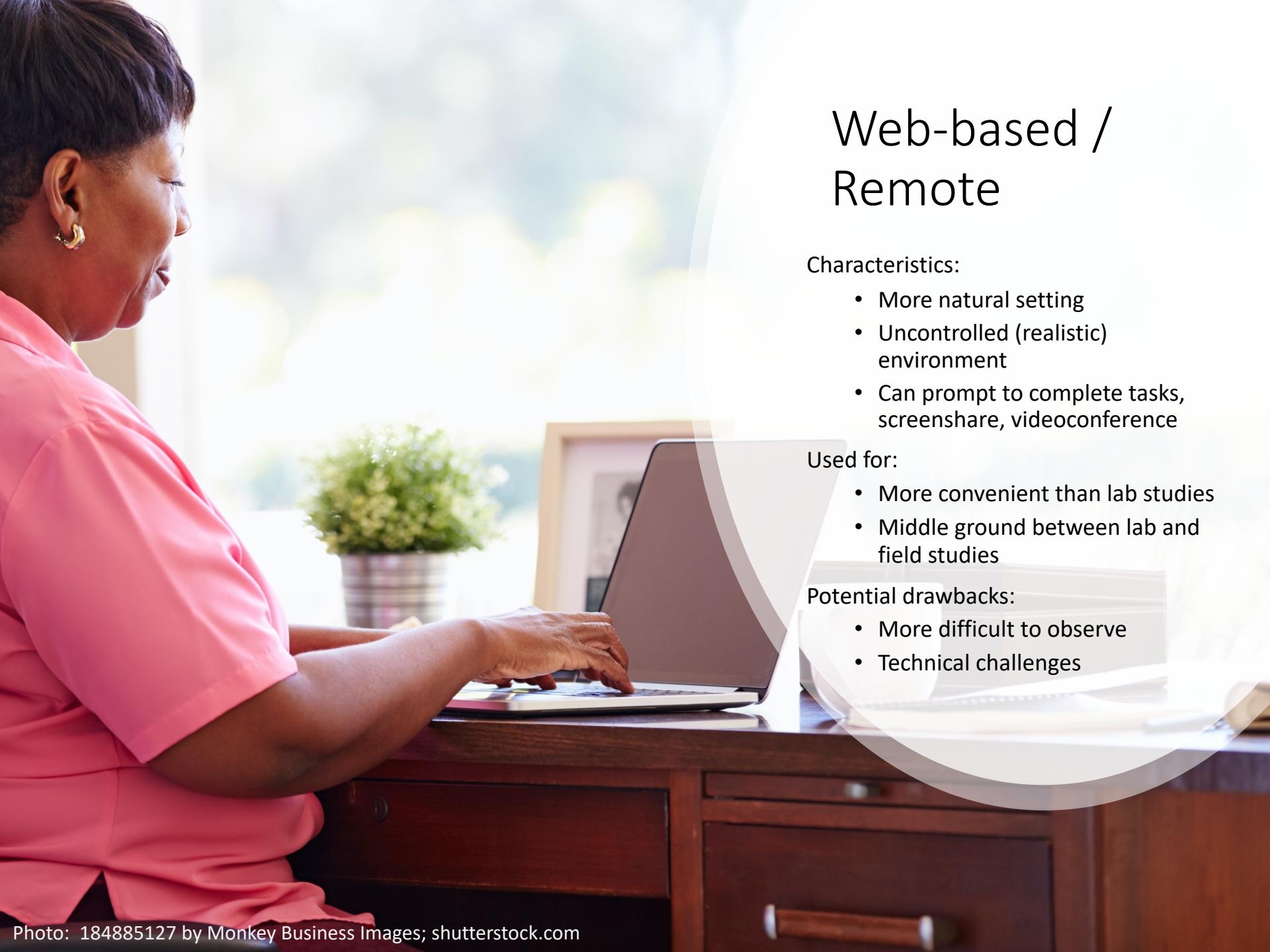
- Evaluate real-world usage
- Gather long-term data

Potential drawbacks:

- Riskier for security, privacy, reliability perspectives
- Influenced by outside factors, unsupervised
- Data much more “messy”



Photo: 1057922507 by Freeograph; shutterstock.com

A photograph of a woman with short dark hair, wearing a pink polo shirt, sitting at a wooden desk and working on a silver laptop. She is looking down at the screen. On the desk next to her is a small potted plant in a brown pot. In the background, there is a window with a view of greenery. A large white circular graphic overlays the top right portion of the slide.

Web-based / Remote

Characteristics:

- More natural setting
- Uncontrolled (realistic) environment
- Can prompt to complete tasks, screenshare, videoconference

Used for:

- More convenient than lab studies
- Middle ground between lab and field studies

Potential drawbacks:

- More difficult to observe
- Technical challenges



Crowd- sourced

Characteristics:

- Fast and cheap data collection through online crowdsourcing

Used for:

- Quickly gathering large amounts of data
- Studies that can be broken into small tasks

Potential drawbacks:

- Data quality/reliability
- Need fully automated study protocol

Triangulation

- Explore the same topic in multiple ways to gather more holistic understanding, increase validity, credibility
- Uses multiple
 - Data sources (people, places, times)
 - Data collection methods
 - Evaluators





Setting up user studies

Usability Testing VS Experiments

Usability testing

- Improve products
- Fewer participants
- Results inform design
- Usually not completely replicable
- Conditions controlled as much as possible
- Procedure planned
- Results reported to developers

Experiments

- Discover knowledge
- Many participants
- Results validated statistically
- Must be replicable
- Strongly controlled conditions, or well-understood tradeoffs
- Experimental design
- Scientific report to scientific community

Usability testing



Goals & questions focus on how well users perform tasks with the product



Comparison of products or prototypes is common



Focus is on time to complete task & number & type of errors



Data collected by video & interaction logging



Testing is central



Satisfaction questionnaires & interviews provide data about users' opinions

Experiments



Predict the relationship between two or more variables.



Independent variable is manipulated by the researcher

E.g.: Type of input: mouse vs stylus
The things you are comparing



Dependent variables are what you are measuring

E.g.: Time, # of errors, preference
How you are comparing them



Validated statistically & replicable.

Different Participant / Between-Subjects Study Design

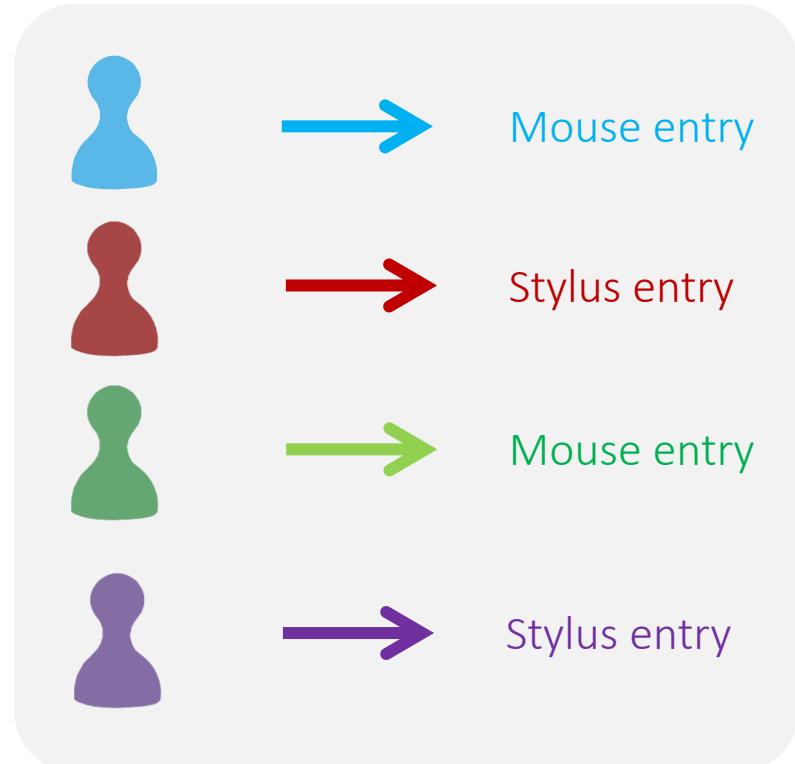
Each participant is in only one test condition

Pros:

- No ordering or training effects

Cons:

- Need more participants,
- Chance of variations between subjects



Same Participant / Within-Subjects Study Design

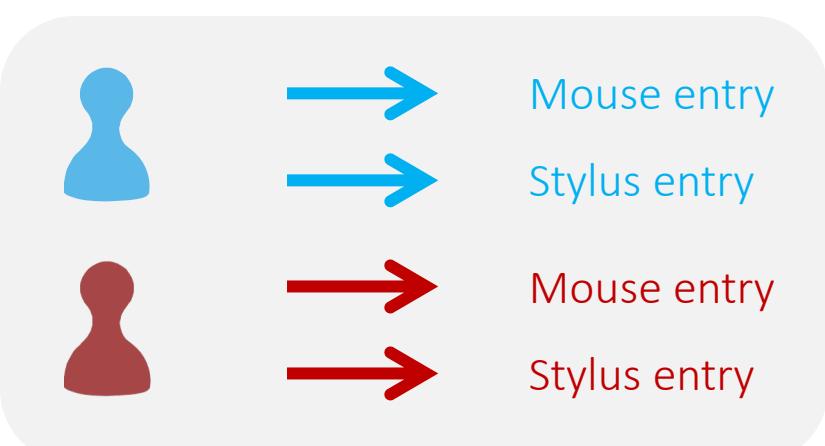
Each participant is in all test conditions

Pros:

- Need fewer participants
- Can see variation in performance across conditions per participant
- Eliminates variation of participants between conditions

Cons:

- Ordering effects
- Training effects



Developing a study protocol



Type of study



Evaluation measures, conditions,
instrumentation



Tasks, scenarios



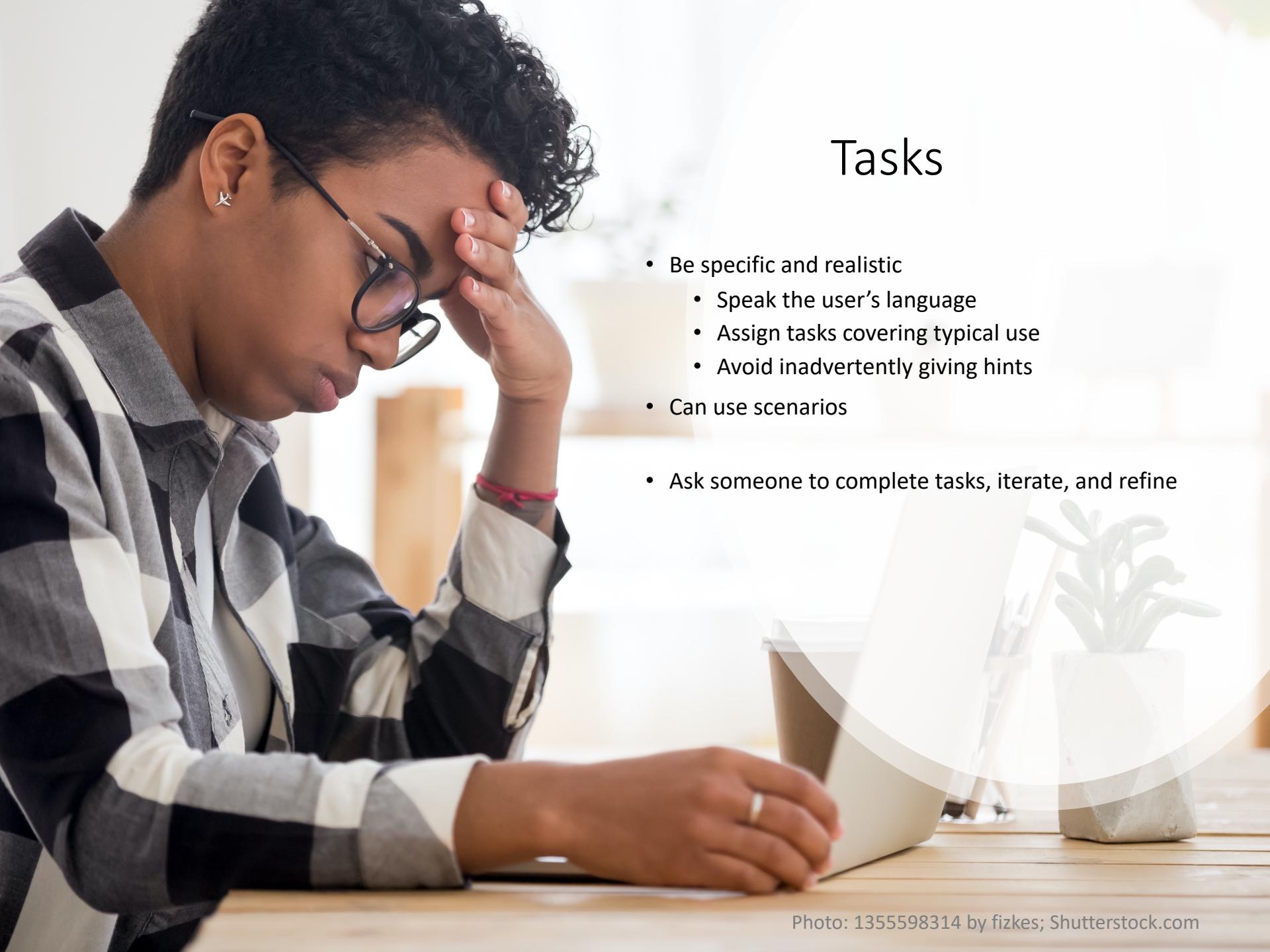
Scripts and instructions



Questionnaires, Interview scripts



Hardware and software

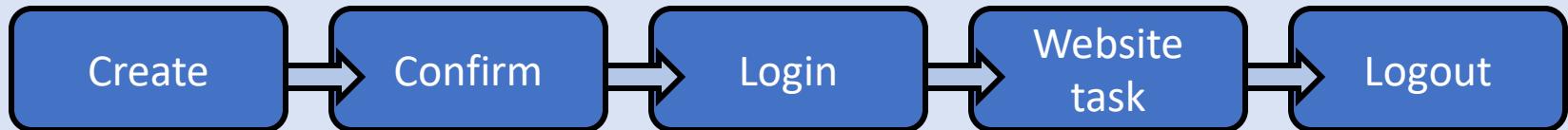
A photograph of a woman with dark curly hair and glasses, wearing a plaid shirt. She is resting her head on her right hand, looking down at a laptop screen. The background is blurred, showing an office environment with a potted plant.

Tasks

- Be specific and realistic
 - Speak the user's language
 - Assign tasks covering typical use
 - Avoid inadvertently giving hints
- Can use scenarios
- Ask someone to complete tasks, iterate, and refine

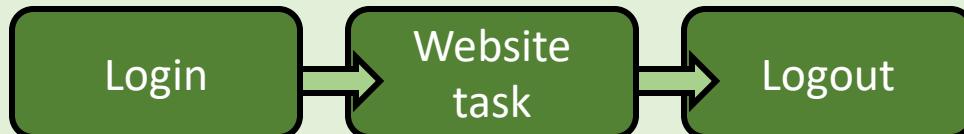
Task descriptions

Password Generation Phase:



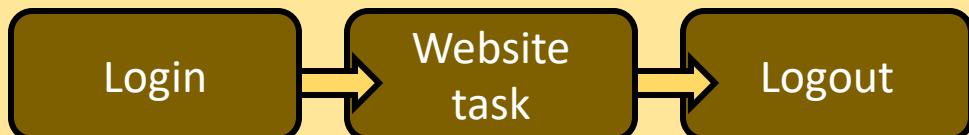
Repeated 3 times, once for each web account

Recall Phase:



Repeated 3 times, once for each web account, in different order

At-Home Phase:



Email sent on day 2, day 4, and day 7

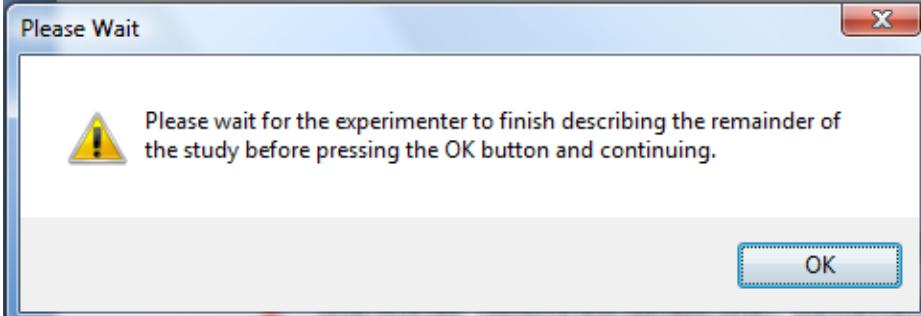
Instructions and Interaction

- Instructions matter, can easily bias user behaviour
 - **Bad:** “comparing our new system to the old system”
 - **Bad:** “This is a program that our start-up company has been working on for 2 years”
 - **Bad:** “We want to see if passwords are too hard to remember”
- Users help evaluate/test the system, users are not being tested
 - You are grateful for any feedback – good or bad!
- Remain neutral, just keep them talking



Software and instrumentation

- Tools for recording user behaviour
 - Screen recording (e.g., Camtasia)
 - Noldus Observer, Morae Observer
 - Eye-tracking, physiological measures
 - Video
- Instrument your prototype
 - Log user actions, timing info
 - Test for accuracy and format of logs!!
 - Include testing-specific prompts and messages
- Plan what kind of analysis you will be doing ahead of time to make sure you have right data



Please wait for the experimenter to finish describing the remainder of the study before pressing the OK button and continuing.

OK

```

# For a given set of clickpoints, tolerance, and image, thi
getHotspotClickpoints = function(dataFrame, tolerance, ima
{
  # Calculates the number of clickpoints around each
  calculateHotspotSize = function(imageClickpoints,
  {
    imageClickpoints$hotspotSize = 0 # Creat
    for(i in 1:nrow(imageClickpoints))
    {
      x = imageClickpoints$x[i]
      y = imageClickpoints$y[i]
      imageClickpoints$hotspotSize[i] = nrow(imageClickpoints[(
      (abs(imageClickpoints$x - x) <= tolerance) &
      (abs(imageClickpoints$y - y) <= tolerance)
      ,]) - 1
    }
    return(imageClickpoints) # Return the "hotspotSize" augmented data frame
  }
  imageClickpoints = function(dataFrame, tolerance, ima
  {
    calculateHotspotSize(imageClickpoints)
  }
  dataFrame
}
  
```

50



Participants

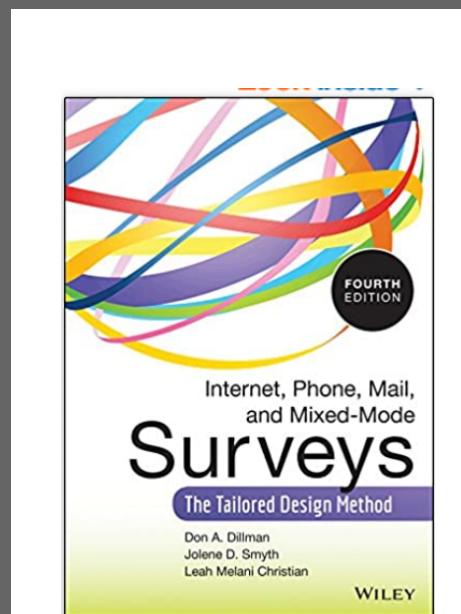
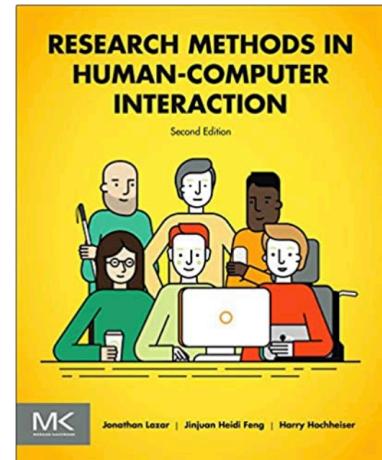
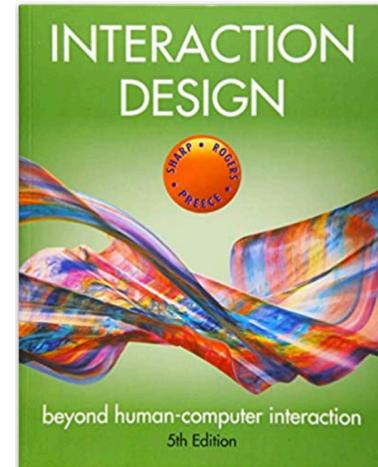
- Representative users
 - Skills, interests, age group, etc.
 - Not your co-workers
 - Get some background info – may help explain or qualify results
- How many?
 - 5-10 for usability testing (availability, cost, scheduling), until no new insight gained
 - 20+ for statistical significance? Experimental studies, surveys
- Compensation
 - What is reasonable?
 - Do not want to influence behaviour

Suggested Resource Books

Sharp, Rogers, & Preece: Interaction Design

Lazar, Feng & Hochheiser: Research Methods in HCI

Dillman, Smyth, & Christian:
Internet, Mail and Mixed Mode
Surveys





A black and white photograph showing six wooden blocks arranged horizontally. Each block has a large, bold, black letter on its top surface. From left to right, the letters spell out "ETHiCS". The blocks are placed on a dark, textured wooden surface. The background is blurred, making the blocks stand out.

ETHiCS



Ethical treatment of participants

- An evaluation can be a distressing experience
 - pressure to perform, errors inevitable
 - feelings of inadequacy
 - competition with other participants
- Participants should always be treated with respect

Before the session

Don't waste user's time	Use pilot tests to debug session Have everything ready ahead of time
Make users comfortable	Emphasize evaluation of the system, not the user Acknowledge that the software may have problems Tell users they can stop at any time
Maintain privacy	Inform user that individual results will be kept completely confidential
Inform user	Explain any monitoring that is being used Answer all user's questions (but avoid bias)
Users must volunteer	User signs an informed consent form Compensate user but do not coerce

During the session

Don't waste user's time

- No unnecessary tasks

Make users comfortable. For example:

- try to give user an early success experience
- keep a relaxed atmosphere
- have snacks, breaks, etc.
- hand out tasks one at a time (don't overwhelm)
- never indicate displeasure with the user's performance
- avoid disruptions
- stop the session if it becomes unpleasant

Maintain privacy. For example:

- do not allow the user's managers (or anyone else) to observe the session



Photo: 1730745325
by Potential Filmmaker;
shutterstock.com



Thank you

Photo: 263790680 by zieusin; shutterstock.com

After the session

- Make the users feel comfortable and valued
 - thank the user for helping you find areas of improvement
- Inform the user
 - answer questions that could have biased the results if disclosed earlier
- Maintain privacy
 - never report results in a way that individual users can be identified
 - only share video/audio tapes with the user's permission

University Research Ethics Board

Any research involving humans must be cleared

Document the study protocol

- strategy, methods, measures, number of subjects, subject recruitment, consent form, etc.

Document the purpose of study

Submitted to the Research Ethics Board

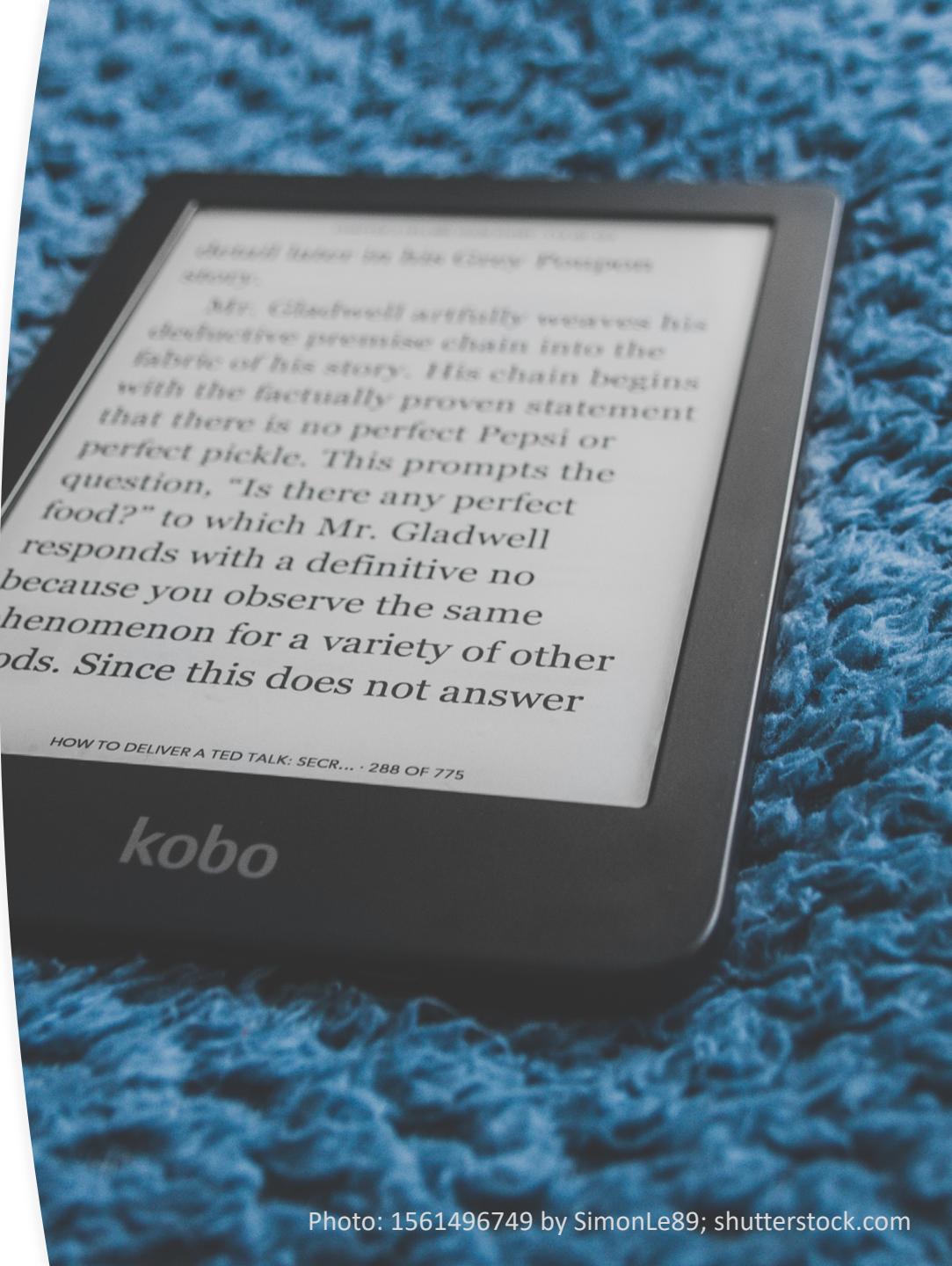
- <https://carleton.ca/researchethics/>
- Reviewed by a committee



Back to Kobo...

Kobo wants to develop a tablet and 'learning-to-read' apps targeted at children and parents. Kobo has e-readers, but hasn't studied how usable these are with children, nor have they looked at competing products.

- What do we want to learn from the user study? What are the evaluation goals?
- What product(s) should we evaluate?
- Who should participants be?
- What activities/tasks are we concerned about?
- What kind of data might we collect?



How do we decide on a type of study?

Work backwards from your goals to figure out what you need to do

1. What are your **overall goals** at this point?
2. What are the **questions** you want to answer?
3. What **kinds of data** do you need to collect?
4. What **type of study** fits these criteria?

1. What are the **study goals**?

1. What are your overall goals at this point in the process?

- may be an iterative process
- may have goals at different levels
 - E.g. – overall: want to design kids' e-reader
 - E.g. – pre-design: are existing apps engaging for kids? Which features currently work/don't work for kids?

2. What are the evaluation questions?

- Current study goals:
 - are existing apps engaging for kids? Which features currently work/don't work for kids?
- What kinds of questions address these goals?
 - How long do kids spend on e-reader before getting bored compared to real books?
 - How do kids feel about using an e-reader?
 - How often do kids get stuck when using the e-reader?
 - What do kids do when they get stuck?
 - What are reading patterns like with an e-reader?
 - How regularly do kids use the e-reader?

3. What kinds of data do you need?

- What kinds of questions address these goals?
 - How long do kids spend on e-reader before getting bored compared to real books? **Timing**
 - How do kids feel about using an e-reader? **Opinion, preference**
 - How often do kids get stuck when using the e-reader? **Counts**
 - What do kids do when they get stuck? **Thought process, personal recollection**
 - What are reading patterns like with an e-reader? **Physiological measures**
 - How regularly do kids use the e-reader? **Real usage pattern, personal recollection, timestamps**

4. What type of study fits these criteria?

- Timing: controlled observation/lab, data logs
- Opinion, preference: interviews, questionnaires
- Counts: controlled observations
- Thought process: Think aloud, observation, interviews
- Personal recollection: Interviews, questionnaires
- Physiological measures: Eye-tracker
- Timestamps: Data logs
- Real usage pattern: Field observations, interviews

Now what?

- On what do you want to focus?
 - Which questions? Which methods?
- Make a specific plan for the user study
 - Type of study (considering time, resources, expertise)
 - Protocol
 - Instruments
 - Data analysis
- Realism? (will results apply in real world?)
- Generalizability? (will results apply to other situations?)

Back to Kobo...

- What tasks could we ask users to do?