CPEN441: user interface design

evaluation I: introduction & ethics, observation & interviews

User evaluation I

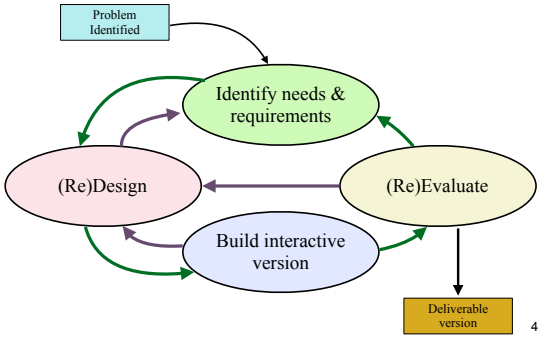
evaluation: observation, interviews, questionnaires

you will use these methods in your project!

evaluation: key things

- evaluation occurs **throughout** the design process
- there are many different evaluation methods - many of which can be used at **any** time during the design process
- evaluation is often equated with user testing that is done at the end of the design process – that is only **one** type of evaluation!

Evaluation at all stages



4

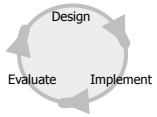
why study users *before* you design?

because you need to know things like:

- **how users do it now**
- **current problems**: inefficiencies, frustrations, lack of or confusing functionality
- **current dependencies**: what parts of the current system are valid, and need to be retained?
- if you have an **approach** for a new design, is it generally likely to solve existing problems?

5

evaluation at various stages



tied to the usability engineering lifecycle

- pre-design
  - investing in new expensive system requires proof of viability
- initial design stages
  - develop and evaluate initial design ideas with the user
- iterative design
  - does system behaviour match the user's task requirements?
  - are there specific problems with the design?
  - can users provide feedback to modify design?
- acceptance testing
  - verify that interface meets expected performance criteria

    ease of learning, usability, user's attitude, performance criteria:

    "A first time user will take 1-3 minutes to learn how to withdraw $50 from the automatic teller"

6

empirical methods of studying users

experiments and quasi-experiments
- observe/measure under controlled conditions

field studies
- observe naturally occurring systems

sample surveys / questionnaires / interviews
- ask people to report on themselves

interpretive methods
- case studies, ethnography, action research, etc.

... and afterward: analysis, modeling, and theory

most common user-based evaluation methods

observation
- passive or participatory, intrusive or unintrusive

interviews
- structured or unstructured

questionnaires
- targeted or open-ended

experiments
- controlled or uncontrolled

what can you expect to learn?

spectrum of data…

qualitative:
- **users tell you** of problems & situations of which they are aware
- **you observe** situations that users may not be fully aware of, due to their immersion
- many approaches to acquire and analyze data
  - Great reference: J. M. Morse. Qualitative nursing research: A contemporary dialogue, Sage, 1991

what can you expect to learn?

spectrum of data…

quantitative:

- measure **task performance** with existing tools / methods: speed, errors, dead-ends, learning curves for novice users, …
- numerical data from questionnaires: # of computers owned, # of email messages received per day, …
- Likert scales of subjective measures
- remember: measures of subjective variables are quantitative data

for both, what you get is influenced by how you ask the question!

characteristics of methods

properties

- intrusive to unintrusive
- formal to informal
- abstract to concrete

criteria

- generalizability
- precision / detail
- realism

multiple methods often required

"triangulation"

intrusive methods

- participant observation
- interviews
- questionnaires (beliefs/attitudes)
- diaries (times and events)
- observation
- "think aloud" protocols
- explicit audio/video recording
- physiological traces

unintrusive methods

- gaze or eye movement traces
- system logs (including Web logs)
- (hidden) observation
- (hidden) audio/video recording
- archives

threats to validity: making sure your data is good?

- **construct** validity
  - are we measuring what we think we are measuring?
    - i.e. complex math questions in English end up measuring English skills, not math
- **internal** validity
  - is there a causal relation between independent & dependent variables?
    - i.e. people work hard because they get paid a lot
- **statistical** validity
  - could the results be a fluke?
    - if we ran the study again, would we get the same results?
- **external** validity
  - do the results generalize beyond test conditions?
    - my study showed that UBC students love my new interface, so will people in Vancouver also like it?
- **ecological** validity
  - do the results generalize to the real-world?
    - i.e. in my driving simulations experiment, participants caused 25% fewer accidents when using driver assist interface
- **face** validity
  - does the experiment subjectively look like it is measuring what it is supposed to?
    - i.e. would questionnaires about historical periods to assess art appreciation seem reasonable?

how do you choose a method(s)?
depends on goals, questions, & constraints

- causality (internal validity), generalizability (external validity) and realism (ecological validity)
- natural vs. artificial setting
- positivist (objective, factual) vs. interpretive approaches
- general principles vs. understanding a specific event
- time, cost, expertise, or resources available
- stage of development at which the evaluation is performed

specific kinds of user study: how-to

you will be doing these for your project

- observation methods
- interviews
- questionnaires & surveys
- Note:
  - already covered two non-user evaluation techniques
    - Heuristic Evaluation and Cognitive Walkthrough
  - one user observation/interview evaluation techniques
    - contextual inquiry

observation

observation

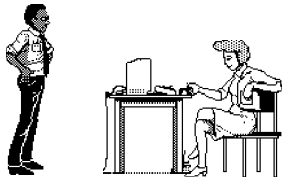
four common approaches:

- simple observation
- think-aloud
- co-discovery learning
- ethnography (covered already)

simple observation

user is given the task (or not), and evaluator just watches the user

problem: no insight into the user's decision process or attitude



think aloud method

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

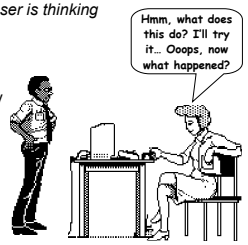
- what they believe is happening
- what they are trying to do
- why they took an action

→ *gives insight into what the user is thinking*

problems

- awkward/uncomfortable for subject (thinking aloud is not normal!)
- "thinking" about it may alter the way people perform their task
- hard to talk when they are concentrating on problem

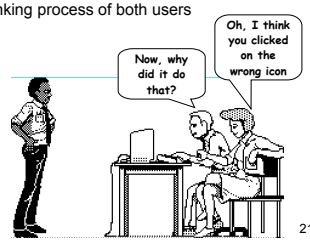
most widely used evaluation method in industry



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
- primitive but cheap
- evaluators record events, interpretations, and extraneous observations
- 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 streams (e.g., interface actions)
- transcription is slow and difficult!

recording observations

video recording:
- can see what a user is doing
  (good to use one camera for screen + one for subject)
- can be intrusive (at least initially)
- analysis can be challenging

companies often build “usability labs” with one-way mirrors, video cams, etc.

analyzing & interpreting observation data

- qualitative data – interpreted to tell a “story”
- qualitative data – categorized, e.g., contextual analysis
- quantitative data – presented as values, tables, charts and graphs; often treated statistically

how do you know which analysis is appropriate?

interviews

querying users via interviews

"conversations with a purpose"

excellent for pursuing specific issues

- more interaction than observation:
  address specific questions of interest
- more flexible than questionnaires:
  probe more deeply on interesting issues as they arise

problems

- accounts are subjective
- time consuming (to conduct and to analyze)
- evaluator can bias the interview
- prone to rationalization of events/thoughts by user
  - *user's reconstruction may be wrong*

planning the interview

general

- what is purpose of interview?
- list of interviewees (breadth vs. depth)
- length of interview & number of sessions
- scheduling interviews (location, times, people)
- will the interview be recorded? (affects the outcome)
  audio, video; transcription

avoid:

- asking long questions
- using compound sentences, i.e. avoid 'and' and 'or'
- using designer jargon
  - OK to use interviewee language
- asking leading questions

… and generally be alert to unconscious biases.

interviews

three main types:

1. open-ended / unstructured
2. semi-structured
3. structured

control & pre-determined questions

other categories (can include types above):

4. group
5. retrospective

unstructured interviews

- most like a conversation, often go into depth
- *open* questions
- exploratory

pros/cons:

+ rich data, things interviewer may not have considered
- easy to go off the rails
- time-consuming & difficult to analyze
- impossible to replicate

*absolute key is for you to **listen** rather than talk:*
practice silence!

structured interviews

- predetermined questions
  (like questionnaire, often with a flowchart)
- *closed* questions
- short, clearly worded questions
- confirmatory

pros/cons:

+ replicable
- potentially important detail can be lost

Better (cheaper) with a questionnaire?

FedUni CLIPP video on YouTube with Professor Simon Cooper and Louise Allen, Nov 17, 2020

semi-structured interviews

in between structured & unstructured

in HCI, un- and semi-structured are the most common

group interviews (focus group)

- 3 – 10 people interviewed at one time
- usually has agenda, but may be structured/unstructured
- skilled moderator critical!
- usually recorded

pros/cons:

+ can accommodate diverse and sensitive issues
+ opinions developed within a social context
+ good way to locate “proto-users”: most articulate, imaginative participants can help later w/participatory design
- some interviewees may dominate
- expensive: usually pay participants + professional moderator

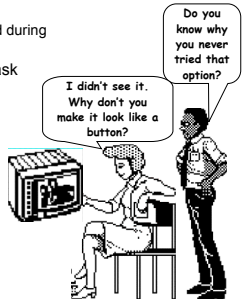
retrospective interview

post-interview to clarify events that occurred during system use:

record what happened, replay it, and ask about it

pros/cons:

+ excellent for grounding a post-test interview
+ avoids erroneous reconstruction
+ users often offer concrete suggestions
- requires a second session



overview: sample exploratory interview

1. explain purpose of the interview
   - time to get acquainted with the interviewee
   - provides understanding and background
2. enumerate activities
   - find out **what** the user does
3. explain work methods
   - find out **how** the user does things (skills and knowledge)
4. trace interconnections
   - determine other people and activities that are related
5. identify performance issues
   - explore current problems and impediments to success

things you uncover during interviews

exceptions

- lots of things people do are not “in the manual”
- many jobs evolve to fit changing circumstances
- much of this is not documented
- many times management does not know about this

domain knowledge

- most people know a lot about their jobs, and those they work with

terminology, common phrases, specific details

- audio recording helps capture this
- video recording helps provide body language
- written notes can provide context, but not always details

summary

- introduction to evaluation
- observation
- interviewing

More resources

- example interview:
  - https://www.coursera.org/lecture/understanding-user-needs/3-1-sample-interview-s13MZ