## Qualitative Evaluation

**CSE333: Introduction to Human-Computer Interaction** 

Jaeyeon Lee

Spring 2023

# Planning user study

## Methods, tasks, and users

#### Collecting data about user performance on predefined tasks

- Data
  - Recordings facial recordings, think-aloud session, ...
  - Logging keystrokes and mouse movements
  - Feedback from questionnaire and/or interview after experiments
- Tasks
  - Searching for information, reading difference typefaces, navigating through different menus
- Performance measures
  - Time and number

## Methods, tasks, and users

Collecting data about user performance on predefined tasks

- Users
  - Users that represent your target user group.
  - Number of participants? It depends.
    - budget, time constraints

## Usability lab





- Testing room
  - The product and recording equipment
  - May be arranged to mimic the real world (office, hotel, ...)
  - Free of the normal sources of distraction (telephone, ...)
- Observation room
  - Observers can see ... through a video or a one-way mirror

## Other settings

- Mobile usability lab
  - A camera for video recording of user's behaviour, a phone for audio recording
  - You can set it up in the real environment

- Remote testing
  - Experimenters help installing app, test online (desktop and mobile interfaces)

## Qualitative Evaluations 1 & 2

**CSE333: Introduction to Human-Computer Interaction** 

Spring 2023

Jaeyeon Lee

### Quantitative vs. Qualitative

- When collecting and analyzing data,
  - Quantitative methods deals with numbers and statistics,
  - Qualitative methods deals with words and meanings.

• Both are important for gaining different kinds of knowledge.

### Quantitative methods

- is expressed in numbers and graphs
- test or confirm theories and assumptions (e.g., hypothesis testing)
- establish generalizable facts about a topic
- include experiments, observations recorded as numbers, and survey with closed-end questions

### Qualitative methods

- is expressed in words
- understand concepts, thoughts, or experiences
- gather in-depth insights on topics that are not well understood
- include interviews with open-ended questions, observations described in words, and literature reviews that explore concepts and theories

### Qualitative methods are used

- To explore the inner experiences of participants
- To explore how meanings are formed and transformed
- To explore areas not yet thoroughly researched
- To discover relevant variables that later can be tested through quantitative forms of research
- To take a holistic and comprehensive approach to the study of phenomena

### Qualitative Methods

#### Query and Critique

- Ask people to provide judgements or feedback on a design
- Excellent for getting initial and continued feedback on a design
- May not identify all problems that will occur
- Some methods are time consuming

#### Observe Users

- Observe users using the system or prototype
- Typically quick and easy to do
- Some methods can reveal what a person is thinking

### Qualitative Methods

#### Query and Critique

- Ask people to provide judgements or feedback on a design
- Excellent for getting initial and continued feedback on a design
- May not identify all problems that will occur
- Some methods are time consuming

#### Observe Users

- Observe users using the system or prototype
- Typically quick and easy to do
- Some methods can reveal what a person is thinking

### Introspection

- Designer/developer tries the system (or prototype) out
  - Does the system "feel right"?
- Pros:
  - Can notice some major problems in early versions during everyday use
- Cons:
  - Not reliable and completely subjective
  - Not valid as the tester is a non-typical user
  - Intuitions and introspection are often wrong
  - Observations may be biased

### Review

- Literature review to support/refute design elements
- Pros:
  - Uses huge prior base of knowledge
  - Saves having to redo established experiments
  - Allows cross-study comparisons
- Cons:
  - Results of a study may not apply to other situations

## Design Critique

- Useful constructive feedback from a design
  - Collaborative deconstructions of what makes a design successful
  - And what makes it fail...
  - Often used in a studio setting

#### • Pros:

- Two-way: designer can provide rationale
- Critic can engage deeply
- Can articulate new design possibilities

#### • Cons:

- Can become abusive
- May not involve real users

### Interviews

- Interview users after they have used the system
- Pros:
  - Excellent for pursuing specific issues
  - Often leads to specific constructive suggestions
- Cons:
  - Accounts are subjective
  - Time consuming
  - Evaluator can easily bias the interview (e.g., did you like the system?)
  - User's reconstruction may be wrong

- Confirmation bias
- Unanswerable question bias
- Social acceptance bias
- Moderator bias

#### Confirmation bias

 Avoid sending signal what you'd like the response to be, or with prior judgement, or your own evaluation.

e.g., Many students prefer this book. What do you think?

- Unanswerable question bias
- Social acceptance bias
- Moderator bias

- Confirmation bias
- Unanswerable question bias
  - Avoid asking about information that can't be answered reliably.
     e.g., that are highly specific, requiring foresight, or others' experiences.

```
"What time did you have dinner 2 weeks ago?"
"How do you think you'd feel after using this for a year?"
"What did your daughter think was the best game?"
```

- Social acceptance bias
- Moderator bias

- Confirmation bias
- Unanswerable question bias
- Social acceptance bias
  - Be careful about socially desirable/undesirable questions.
     e.g., requiring answers that people may be reluctant to be honest

```
"When was the last time you flossed?"
"Have you ever committed a crime?"
```

Moderator bias

- Confirmation bias
- Unanswerable question bias
- Social acceptance bias
- Moderator bias
  - They may want to please you rather than being honest.

## Retrospective Testing

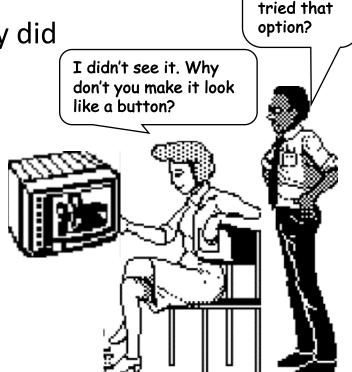
- Interview to clarify events that occurred during system use
  - Perform an observational test
  - Create a video record of it
  - Have users view the video and comment on what they did

#### • Pros:

- Excellent for grounding a post-test interview
- Avoids erroneous reconstruction
- Users often offer concrete suggestions

#### • Cons:

- Time-consuming
- Users may forget context



Do you

know why

you never

## Questionnaires & Surveys

- Set of questions user answer on their own
- Pros:
  - Maintenance / analysis is easy
  - Can reach a wide subject group (e.g. mail / web)
  - Does not require presence of evaluator
  - Results can be quantified

#### • Cons:

- Preparation can be time-consuming
- Only as good as the questions asked
- Users may misunderstand, or not put in effort

### Qualitative Research Methods

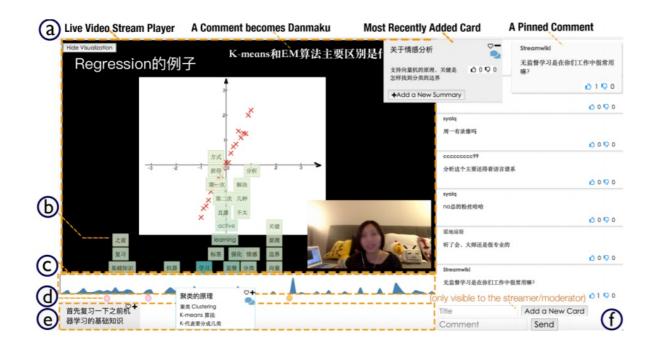
- Query and Critique
  - Ask people to provide judgements or feedback on a design
  - Excellent for getting initial and continued feedback on a design
  - May not identify all problems that will occur
  - Some methods are time consuming

#### Observe Users

- Observe users using the system or prototype
- Typically, quick and easy to do
- Some methods can reveal what a person is thinking

## Conceptual Model Extraction

- Show the user static images of:
  - Paper prototype, screen snapshots, actual system screens during use
- Have the user try to explain
  - What all elements are
  - What they would do to perform a particular task



### Conceptual Model Extraction

- Initial vs formative conceptual models
  - Initial: how person perceives a screen the very first time it is viewed
  - Formative: how person perceives a screen after the system has been used for a while
- This approach is:
  - Excellent for extracting a novice's understanding of system
  - Poor for examining system exploration and learning
  - Requires active intervention by evaluator, which can get in the way

### Direct Observation

- Evaluator observes and records users interacting with design/system
  - In the Lab: User asked to complete a set of pre-determined tasks
  - In the Field: User goes through normal duties
- Excellent at identifying gross design/interface problems
- Validity/reliability depends on how controlled the environment is
- Three typical approaches:
  - Silent Observer, Think-Aloud, Constructive Interaction

### Direct Observation: Silent Observer

- User is given the task, and evaluator just watches the user
- Problem
  - Does not give insight into the user's decision process or attitude
  - User's can become frustrated if stuck on a task



### Direct Observation: Think-Aloud

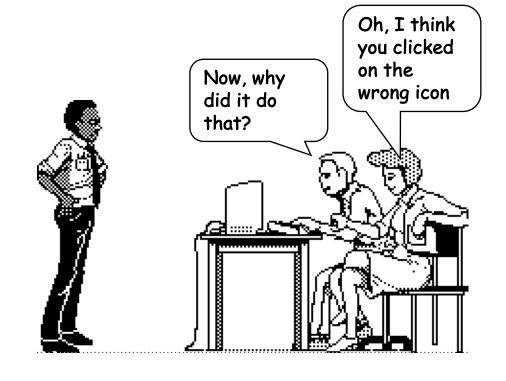
- Subjects are asked to say what they are thinking/doing
  - What they believe is happening
  - What they are trying to do
  - Why they took an action
- Gives insight into what the user is thinking
- Problems
  - Awkward/uncomfortable for subject (thinking aloud is not normal!)
  - "Thinking" about it may alter the way people perform their task
  - Hard to talk when they are concentrating on problem
- Most widely used evaluation method in industry

Hmm, what does this do? I'll try it... Ooops, now what happened?



### Direct Observation: Constructive Interaction

- Two people work together on a task
  - Normal conversation between the two users is monitored
    - Removes awkwardness of think-aloud
  - Variant: Co-discovery learning
    - Use semi-knowledgeable "coach" and naive subject together
    - Make naive subject use the interface
  - Results in:
    - Naive subject asking questions
    - Semi-knowledgeable coach responding
    - Provides insights into thinking process of both beginner and intermediate users



## Direct Observation: Recording Observations

#### Paper and pencil

- Evaluators record events, interpretations, and extraneous observations
- Hard to get detail (writing is slow)

#### Audio recording

- Good for recording talk produced by thinking aloud/constructive interaction
- Hard to tie into user actions (i.e. what they are doing on the screen)

#### Video recording

- Can see and hear what a user is doing
- One camera for screen, another for subject (picture in picture)
- Can be intrusive during initial period of use
- Companies often build "usability labs" with one-way mirrors, video cams, etc.

#### Software Logs

- Log and time-stamp key events that can be later analyzed
- May require modification of the software being tested

### Continuous Evaluation

- Case Studies, Field Studies
  - Careful study of "system usage" at the site
  - External observer monitors behavior

#### • Pros:

- Good for seeing "real life" use
- Actual tasks, actual users, actual environments

#### • Cons:

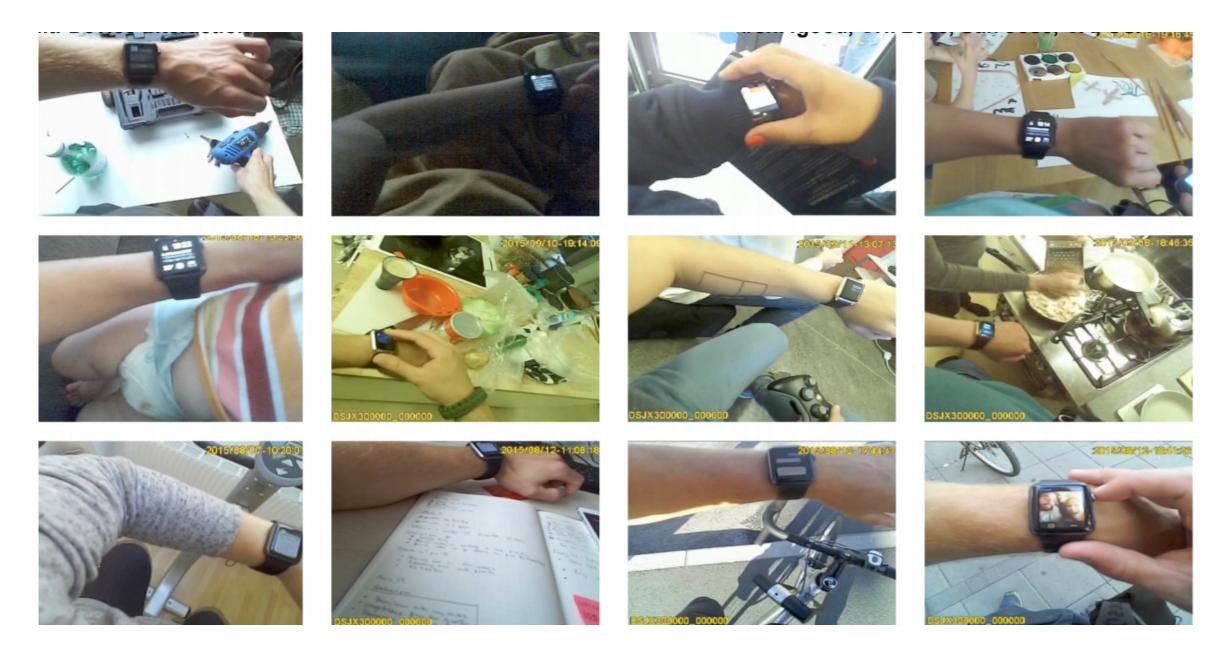
- Time consuming
- Low sample size
- Can only test situations that arise



Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. Smartwatch in vivo. CHI '16



Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. Smartwatch in vivo. CHI '16



Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. Smartwatch in vivo. CHI '16

### Getting Situated and Contextualized Data

- To understand participants' experiences natural settings, in their words without having an observer
- It can be collected through
  - Diary Study
  - Experience Sampling

### Diary Study

- Participant-driven data collection: participants write a diary about their experience.
- Self-report at regular intervals, or when an event of interest occurs
- Participants create logs of activities answering given questions.

# Diary Study

#### • Pros:

- Non-intrusive, non-interrupting
- Potential for detailed answers

#### • Cons:

- Rely on participants' motivation, effort, and memory
- May lose many events if participants forget to log

### Experience Sampling

- Experimenter-driven data collection: participants are notified when they need to report their experiences.
  - E.g., using a beeper, phone messages, or custom apps.
- Though sampling
  - Random paging to determine "mental content"
  - Experience sampling: Likert scales or multiple-choice questions to assess the context of experience
  - Descriptive Experience sampling: qualitative and open-ended questions.

### Experience Sampling

#### Pros

- Less likely to misremember behavior
- Collect data anytime with no observer needed

#### Cons

- Interrupting and inconvenient
- May miss important experiences
- Potential for post-hoc rationalization

### Qualitative Research Methods

### Query and Critique

- Ask people to provide judgements or feedback on a design
- Excellent for getting initial and continued feedback on a design
- May not identify all problems that will occur
- Some methods are time consuming

#### Observe Users

- Observe users using the system or prototype
- Typically, quick and easy to do
- Some methods can reveal what a person is thinking

# Qualitative Research Methods: Query and Critique

- Introspection
- Review
- Design critique
- Interviews
- Retrospective Testing
- Questionnaires and Surveys

# Qualitative Research Methods: Observing users

- Conceptual model extraction
- Direct observation: Silent observer
- Direct observation: Think-Aloud
- Direct observation: Constructive Interaction
- Continuous evaluation
- Diary study
- Experience Sampling

# Interviews and Focus Groups

How to conduct an interview?

### Getting direct feedback from users

- Ask the users!
- **Surveys** can be very useful in this regards. But they are limiting as respondents only answer questions that are asked, long and written questions are often unanswered.
  - → broad but not deep
- Go deep but not broad: Direct conversations as tools for data collection
  - Interviews: One at a time
  - Focus Groups: Many at a time

### Interviews and Focus Groups

#### Pros

- Rich, attitudinal data
- Can gain in-depth responses about experiences, attitudes, and knowledge
- Flexibility

#### • Cons

- Need skill to manage
  - Moving conversation along, eliciting meaningful responses, interpreting subtle cues, etc.
- Time/resource intensive
- Recall problems

### Applications of interviews

- HCI researchers use interviews and focus groups to help build an understanding of the needs, practices, concerns, preferences, and attitudes of the current or future users.
  - Can be used in any phase
- Initial exploration
- Requirements gathering
- Evaluation and Subjective Reactions

### Interviews for initial exploration

 When you're embarking on a new project involving the design of novel tools for unfamiliar users and needs.

- You might sit down with various potential users to understand their potential goals and needs:
  - What would they like to do with their pictures and videos?
  - How do these artifacts play into their daily lives?
  - Where and when are they most likely to reach for a picture or video?
  - How do they use these records to tell stories about their lives?
  - > not focused on specific questions

### Interviews for requirements gathering

- During the process of gathering requirements for the design of a new tool
- Interviews at this early stage are likely to be broad:
  - What are the user's goals?
  - How are they being met by current tools (if any are available)?
  - What do users want to do that they are currently unable to do?
  - What are the frustrations?
  - →open-ended view of the possibilities
- You might also ask your participants to describe things that they'd like to do, regardless of the feasibility with current software

# Interviews for evaluation and subjective reactions

- Interviews at this stage focus on specific questions to capture the reactions of various users
  - Do you find this interface easy to use?
  - Do you understand the menus, icons, and language?
  - Are you able to complete the comparisons that you want to do? If not, where do you have trouble?
  - Are any parts of the interface particularly useful or helpful?
  - How does this interface compare to your current tool?
  - What (if anything) would you like to change about this tool?
- Additional questions may be based on existing usability testing questionnaires: QUIS, SUS, etc. (will be covered in Quantitative methods)

### Who to interview

- Beyond users Stakeholders
  - Anyone who may be affected the system use
- Interview representatives of different groups from different perspectives
- Key informants: particularly useful/informative individuals

### Interview strategy: how much structure?

### Fully Structured

- Stick with the script
- Good for comparison across individuals

### Semi-Structured

- Pre-specified questions
- Starting point for discussion
- Digression is ok

### Unstructured

- Initial question
- possible list of topics
- But no pre-defined script

### Interview strategy: how much structure?

Fully Structured

Semi-Structured

Unstructured

Less structure: greater difficulty in conducting and interpreting interview

#### **But**

More opportunity for insight

### Un- or Semi-structured Interviews

- Often, questions are generated in response to participant comments
- Can be hard to do this well
- Start with more highly-structured interviews

### Focused & contextual interviews

- Go beyond asking questions
  - to understand not just explicit understanding but also to find implicit practices

- Ask participant to
  - Demonstrate use of technology, not just explaining how they do
  - Show artifacts (papers, photos, etc.)
  - React to "probes" props or prototypes designed to elicit reaction

### Interviews vs. focus groups

- Interviews take time
  - Often 1 hour or more/response
  - Several hours for analyzing notes
  - May be done by a researcher
- Focus groups
  - More people in less time
  - Around 5-7 participants, or up to 8-12 people.
  - May need multiple researchers to keep the conversation going, monitor the clock, walk through the list of questions, all the while collecting the data.

### Focus groups: pros and cons

#### Pros

- Broad range of viewpoints and insights
- Each group will likely have at least one person who will stimulate others to talk

#### Cons

- Hard to manage group dynamics
- Generally, can't be fully structured
- May need to ask fewer questions
- Participant selection can be challenging

### Other guidelines

• Simple questions, no jargon

- Avoid compound questions with multiple parts
  - Not "What were the strengths and weaknesses of the menu layout and the toolbar?"
  - Ask two separate questions instead.
  - Or four
- Avoid judgmental phrasing or tone
  - Possible bias

### Preparing for the Interview

- Pilot test with colleagues and participants
  - Good for logistics and for confidence
- Write an interview guide listing what to do and when
  - Particularly good if multiple researchers are involved
- Logistical backups
  - Batteries for laptop, audio recorder, extra paper, etc.

### Notes

 Audio and video recordings are fine. But paper notes are still important.

- Record insights, non-verbal responses, etc.
  - Try to record what you can, but don't do so at the expense of listening
  - This notes save time when you later analyze the recordings!

• Summarize written notes as soon as possible after the interview before you forget.

### Recordings

- Complete, but expensive
- Transcription can take many hours
- Video is tricky, but gets useful information
- Consider audio + still pictures
- You may also take screenshots/screen recordings under their permission
- Respect privacy and anonymity
- Have a consistent policy for comments made after the notebook is away and the recorder is off (be professional!)

### During the Interview

- You're the host:
  - Build rapport
  - Be friendly, respectful, nonjudgmental
  - Listen carefully

#### Outline

- Briefly introduce research goals
- Complete paperwork (informed consent)
- Simple questions first, hard questions later

### During the Interview

- Be flexible
  - If your interview is not fully structured
- But try to keep things on track
- Explain why you are asking each question
- Define terms, avoid jargon
- Ask for clarification and elaboration

### Closing It Out

- Ask for any final comments
- Provide more detail about research goals
- Brief summary of findings
- Turn off recording devices
  - Interviewees might make additional useful comments
  - Ask before including these comments in analyses
- Say "thanks!"
- Reflect and summarize notes immediately

### Phone or Online Interviews

- Phone, video calls, web chat, email, conference calls
- Pros
  - Easy, inexpensive
  - Reach more/geographically distant people with less effort
  - For interviews involving sensitive issues, good for anonymity
  - Participants are in their own environment so are more relaxed
  - Potentially powerful to capture screen, audio

#### • Cons

- Lack of face-to-face contact
- Fewer non-verbal cues
- Pacing can be harder
- Group discussion can be challenging

### Data analysis

- Do it as *soon* as possible
  - Sit down and summarize notes immediately afterward
- Avoid "cherry-picking"
  - Capture both positive and negative points
- Fully-structured, closed-ended: tabulate answers
- Open-ended questions require <u>qualitative coding</u>
  - Transcribe audio
  - Summarize note

### Qualitative analysis

- Content analysis
  - frequency of terms, patterns in the notes
- Categorization
  - Affinity Diagrams
- Multiple analyses can increase validity

# Analyzing Test Data

- A priori coding
  - Identify coding categories first
  - Coding
  - Reliability check
- Emergent coding
  - Multiple coders identify coding categories based on subset of data
  - Consolidate category list
  - Code a subset of data
  - Reliability check
  - Repeat the process until satisfactory result is met
  - Code the rest of the data

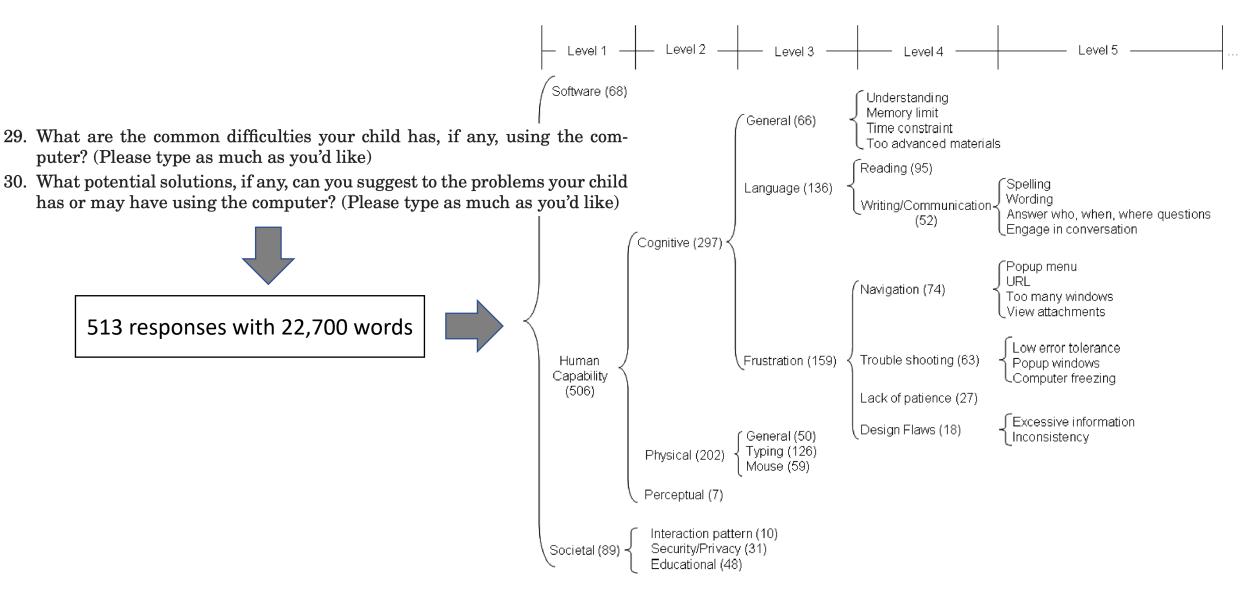


Fig. 1. Summary of difficulties experienced by children with DS (numbers in parentheses represent the number of children whose parents reported the difficulty).

# Coding the text

Look for key items

Statement	Examples
Objectives	Use computers for educational purposes
Actions	Enter a password, chat online
Outcomes	Success or failure, whether the objective is achieved
Consequences	Files unintentionally deleted, a specific application abandoned
Causes	Limited memory, dated equipment
Contexts	User is computer savvy, user works with classified information
Strategies	Avoid specific tasks, multimodal interaction

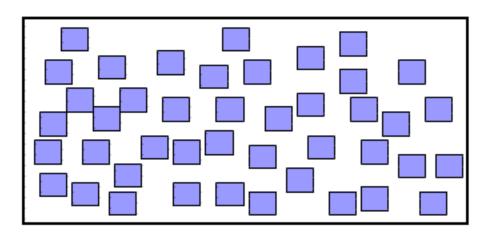
Table 11.2 What to look for while coding.

### Coding the text

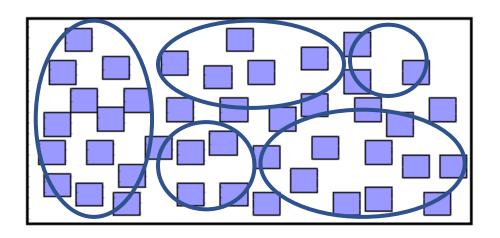
- Ask questions about the data
- Making comparisons
  - Between different coding category
  - Between different participant group
  - Between existing data and previous literature

- Common technique used in qualitative analysis
- Organize individual ideas and insights into a hierarchy showing common structures and themes.
- Groups are not predefined but emerged from the data.

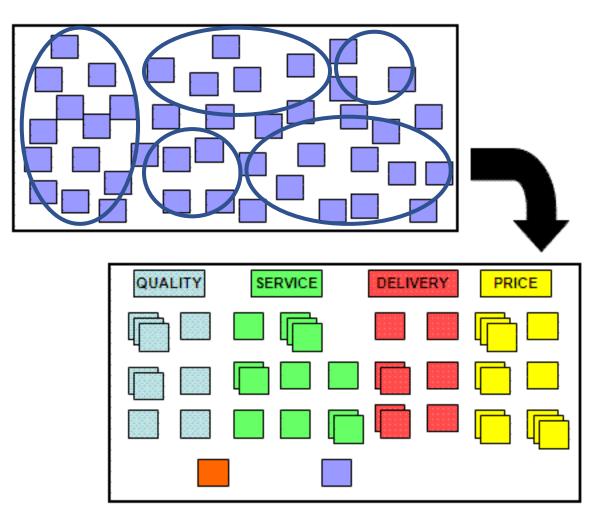
- Start by generating sticky notes
  - Observations and comments



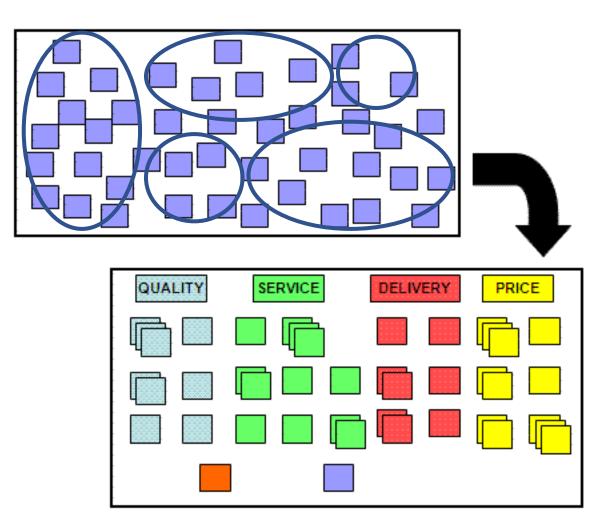
- Start by generating sticky notes
- Organize notes in groups
  - Find similar items and organize in groups
  - Should be done by multiple researchers, but no discussion



- Start by generating sticky notes
- Organize notes in groups
- Sort Notes into top-level categories



- Start by generating sticky notes
- Organize notes in groups
- Sort Notes into top-level categories
- Sort Notes into subcategories



#### Prototype C Feedback

ike that i

an easily

arn more

There's no

titles on the

form

questions

what a farmlet is

from a personal

and enviro.

perspective would

be awesome



Kanban boards (e.g., Trello) are also great for this!

### Reporting results

- Be as specific as possible
  - Not "most respondents"
  - Instead, "7 out of 10 respondents"
- Use quotes or paraphrases from respondents
  - But don't use participant name
  - Use identifiers (e.g., P3) or pseudonyms (pen-name)
- Categories can be a good way, too
  - Organize findings in an organized way

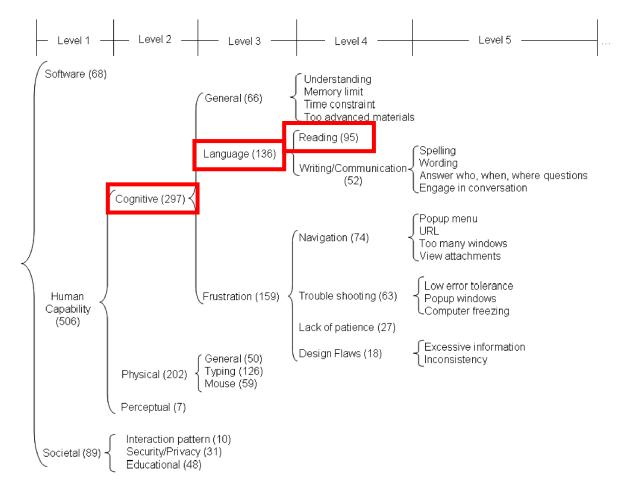


Fig. 1. Summary of difficulties experienced by children with DS (numbers in parentheses represent the number of children whose parents reported the difficulty).

#### 4. COGNITIVE LIMITATIONS

Three subcategories were listed under cognitive difficulties. They are general cognitive difficulties, language difficulties, and difficulties caused by frustration. Language- and frustration-related difficulties were very frequently stated with approximately one out of four respondents reporting difficulties under either of the two categories. 11.8% of the respondents stated difficulties due to general cognitive limitations.

#### 4.1 Language Difficulties

Language difficulties can be further divided into two categories: reading difficulty and writing and communication difficulty. 16.9% of the respondents reported reading difficulties. 9.3% of the respondents reported writing and communication difficulties.

4.1.1 Reading Difficulties. Reading problems are not exclusive to children with DS. Neurotypical children also need time to build their reading skills and experience similar problems (e.g., Chall [1983], Braunger and Lewis [2005], and Mateos et al. [2008]). However, children with DS experience more substantial problems with reading and these problems persist as they mature (e.g., Cardoso-Martins et al. [2008]). Although children with DS demonstrate stronger reading abilities than language and number abilities, their reading

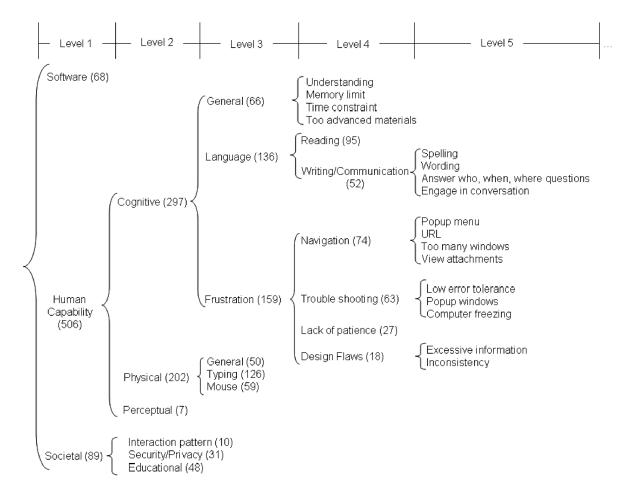


Fig. 1. Summary of difficulties experienced by children with DS (numbers in parentheses represent the number of children whose parents reported the difficulty).

#### 4. COGNITIVE LIMITATIONS

Three subcategories were listed under cognitive difficulties. They are general cognitive difficulties, language difficulties, and difficulties caused by frustration. Language- and frustration-related difficulties were very frequently stated with approximately one out of four respondents reporting difficulties under either of the two categories. 11.8% of the respondents stated difficulties due to general cognitive limitations.

#### 4.1 Language Difficulties

Language difficulties can be further divided into two categories: reading difficulty and writing and communication difficulty. 16.9% of the respondents reported reading difficulties. 9.3% of the respondents reported writing and communication difficulties.

4.1.1 Reading Difficulties. Reading problems are not exclusive to children with DS. Neurotypical children also need time to build their reading skills and experience similar problems (e.g., Chall [1983], Braunger and Lewis [2005], and Mateos et al. [2008]). However, children with DS experience more substantial problems with reading and these problems persist as they mature (e.g., Cardoso-Martins et al. [2008]). Although children with DS demonstrate stronger reading abilities than language and number abilities, their reading

Reading problems make it difficult to retrieve information on the Internet and cause frustration among the children.

"[His difficulty is] understanding how to find things (search, select links to click < overwhelming number of results>, find what he needs on the page) on the Internet; overwhelming number of words on a page, has a hard time following the words... difficulty understanding some directions; frustration leading to giving up... he would benefit from programs which offer auditory instructions rather than requiring reading." (13)

Limited reading skills make the children highly vulnerable to errors and unexpected outputs.

"[His difficulty] is in problem solving...if computer isn't "working" like he wants it to, he gets frustrated. He can't read the error messages so needs someone always nearby." (7)

Lack of reading skills also makes it difficult for children to take full advantage of computers and the Internet as a communication tool.

"[Her difficulty is] lack of reading skills, lack of socialization that enable her to chat, . . . she uses sign language, pecs (picture exchange communication system [Bondy and Frost 2001]), some speech to communicate and that needs to be incorporated into the programs." (14)

The age of the child (minimum information)