Comparative Structured Observation

a qualitative method for capturing comparative user insights to advance novel design concepts

CHI 2025

Wendy E. Mackay
Joanna McGrenere

CHI'25 Course
Yokohama, Japan

Wendy E. Mackay
Inria & Université Paris-Saclay

Psychology ⟶ HCI

Joanna McGrenere
University of British Columbia

HCI ⟵ Computer Science

Schedule

Who are we?

What is Comparative Structured Observation ?

How to create a CSO protocol?

Group activity: Create your own CSO study

What makes a good CSO study?

What is next?

90 minutes

9:00	Introduction
9:05	Comparative Structured Observation
9:15	Walk through an example
9:25	Create a CSO protocol (group exercise)
10:00	Protocol critiques
10:25	Final Discussion and Wrap-up

What is Comparative Structured Observation?

When we create a new design concept

When we create a new design concept

How do we …

assess the **design direction**?

When we create a new design concept

How do we …

assess the **design direction**?

or

advance the **design concept**?

HCI borrows from
different research disciplines

Social Scientists

seek to understand human behavior

Including:
- **quantitative** measures of performance
  (controlled experiments)
- **qualitative** measures of behavior
  (observation and interviews)

HCI uses both …
but for a different purpose

We want to understand human behavior:
to generate “implications for design”

How do HCI Researchers evaluate their results?

Borrow social & natural science methods

run **controlled experiments** to test causal hypotheses related to performance

conduct **interviews** to discover user preferences, e.g. Likert-type questions

run **usability studies** to ensure that the system is “easy to use”

How do HCI Researchers ensure a “good” design?

Few rigorous qualitative design methods

Experiments are designed to detect
cause-and-effect relationships
not assess the quality of the design

Emphasis on **quantitative** measures
require testable hypotheses

What if you want to run an experiment …
but **do not have a hypothesis**?

“Get the design right”

S. Greenberg & W. Buxton (CHI 2008)
Usability Evaluation Considered Harmful (Some of the Time)

“Get the design right”

“Get the right design”

S. Greenberg & W. Buxton (CHI 2008)
Usability Evaluation Considered Harmful (Some of the Time)

How do HCI Researchers ensure a “good” design?

How can we take advantage of
the user’s **reflections** on their **experiences**

to help us **assess incomplete designs**
in a rigorous way?

What can we borrow from experiments
not to discover causal relationships

but to benefit from their structure
to maximize what we learn from users?

HCI researchers need

Citable method name
Clearly defined goals
Clear guidelines
Best practices
Success assessment

Comparative Structured Observation

Comparative Structured Observation

Mixed Qualitative Method

for Gathering Observational Data

Key goals include:

Gather insights from users about a
novel design: is it on the right track?

Identify trade-offs across design variants
compared to the status quo
or among alternative design variants

Explore new design directions,
push the limits of the design or
identify rare or extreme situations

Comparative Structured Observation

Mixed Qualitative Method

for Gathering Observational Data

Key goals include:

Gather insights from users about a novel design: is it on the right track?

Identify trade-offs across design variants compared to the status quo or among alternative design variants

Explore new design directions, push the limits of the design or identify rare or extreme situations

What it is NOT:

An experiment to determine if a particular feature improves performance

An experiment to test user interface details

An open-ended field study to uncover underlying theory about users

Comparative Structured Observation

Generate actionable design implications

Most appropriate for mid-phase design: grounded design ‘hunches’ or ‘intuitions’

Create activities that highlight common but also rare or interesting issues

Focus on gathering new design insights

Comparative Structured Observation

W.E. Mackay and J. McGrenere
Comparative Structured Observation
ACM/TOCHI (January) 2025

Comparative Structured Observation

An interventionist, qualitative method
for assessing and advancing a design concept

Researchers **observe** participants
as they **compare** and reflect on their experiences

with selected design variants
structured according to experimental design principles

Comparative Structured Observation

Comparison

Compare a new design to the **status quo** or other **design variants**

Structure

Structure participant activities according to established **experiment design principles**

Observation

Researcher observes participant activities

Participants reflect upon their experiences

Comparative Structured Observation

Design concept

- Derived from formative studies
- Ecologically valid activities
- Mid-phase interactive prototypes

Data collection

- Capture participant behavior
- Record qualitative reflections
- Qualitative 1º, quantitative 2º

Data analysis

- Established qualitative methods

Results

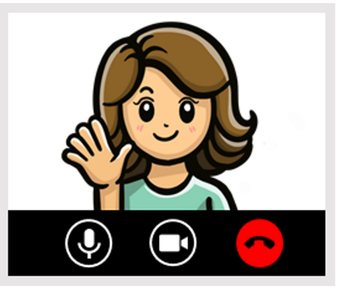
- Advance design concept(s)

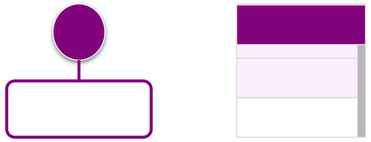
How to create a CSO protocol?

Example:
HelpCall

HelpCall: Design concept

Older Adult's screen + control are auto-shared to the helper





Augmented steps display (ASD)

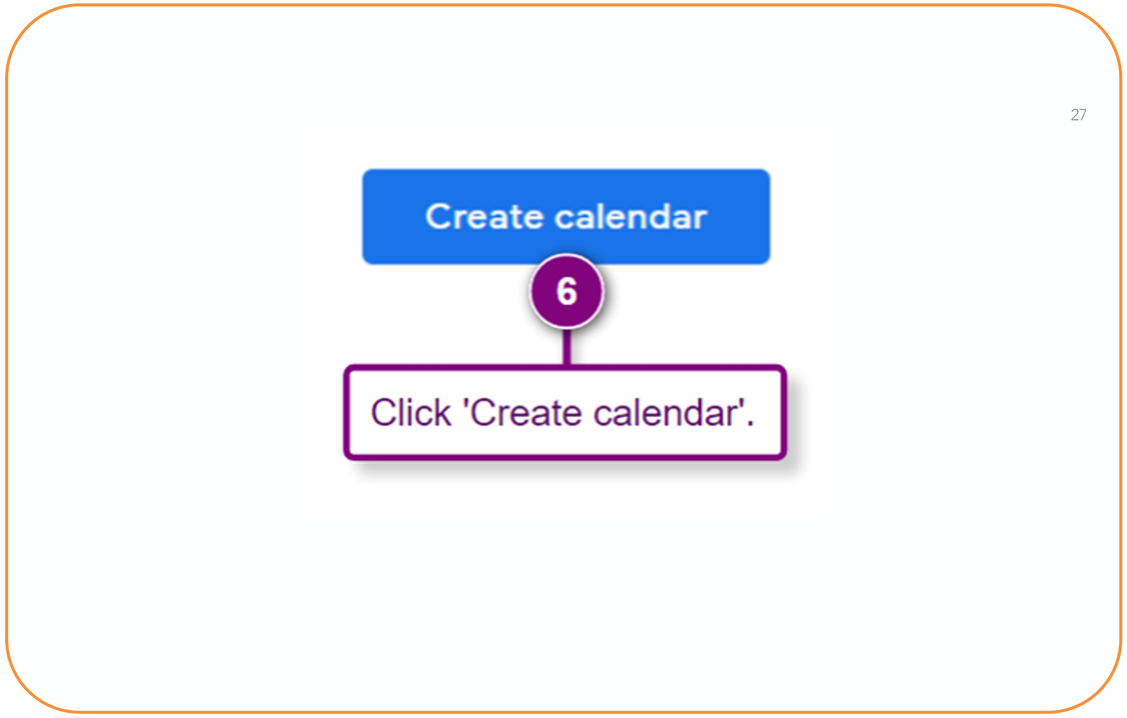
Augmented video-mediated communication

for assisting older adults in learning software tasks

Augmented steps display:

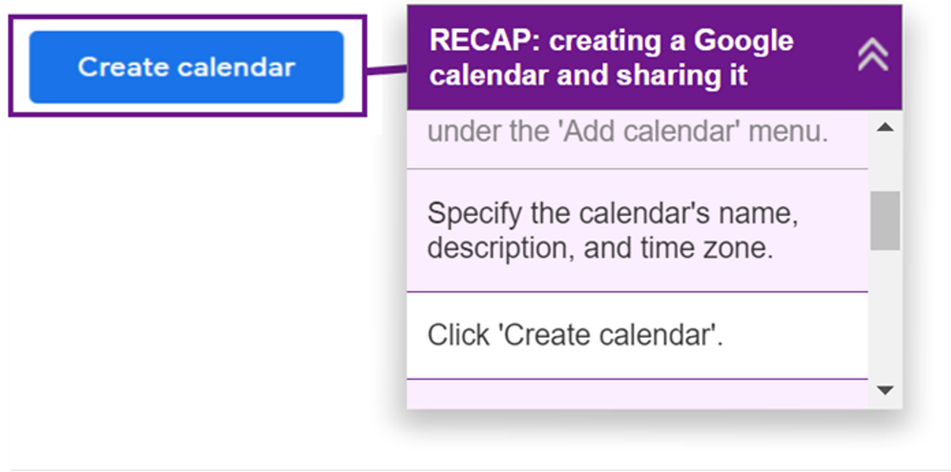
Design variant #1: Tooltip

Demo (helper in-control, learner's perspectives)



Two design candidates

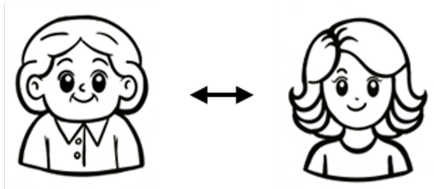
Design variant #2: List



HelpCall: Intended flow

Comparative Structured Observation

Learner-helper synchronous session



Learner's asynchronous re-access

Round 1
Helper demonstration

Round 2
Helper-supported practice

Round 3+
No-helper attempt(s)

14 older adult (65+) participants with computer & video call experience

HelpCall

Design Concept

How to help older adults get tech help?

Family members demonstrate steps
over a video conferencing link

HelpCall
Captures, displays and saves steps
from live demonstrations from family

Two design variants:
Tooltip
List

Is CSO appropriate?

HelpCall

Clearly defined concept?	☑ Family members demonstrate tech features to older adults via video calls
Based on formative research?	☑ Literature, cognitive walkthrough
Clear user population?	☑ Older adults
Interactive prototype?	☑ Medium fidelity interactive prototype task mockup + Wizard-of-Oz
Design variants?	☑ Tooltip, List + Status Quo
Access to users?	☑ In home, simulated video call

Study design

HelpCall

Grounded activities?	☑ Six learning tasks beginner and intermediate levels
Equivalent activities?	☑ Equivalent but not fully isomorphic
Specify conditions?	☑ Three conditions
Task assignment?	☑ Three tasks per participant
Counterbalancing?	☑ Status Quo then alternate Tooltip, List
Setup?	☑ Older adults in home setting (or lab) with simulated video call

Design the protocol

Arrival in participant's home

Ask participant to perform

three different tasks

with each design variant

Semi-structured comparison interview

Final debriefing

HelpCall

- [x] Informed consent
  Study briefing
- [x] Task 1 with Status Quo + survey
- [x] Task 2 with [Tooltip]* + survey
- [x] Task 3 with [List]* + survey
- [x] Comparison interview
- [x] Debriefing

* Order varies per participant

Data Collection

HelpCall

Type of data collected?	☑ Qualitative first Quantitative second
Researcher observation?	☑ In-person observation Screen recording Interaction log
Participant reflection?	☑ Interview Comparison questionaire
Analysis	☑ Mixed thematic analysis

Group work

Groups of 3 or 4

1. Pick a design concept (5 minutes max)
   your design or this one 

2. Go through the checklist:
   Is CSO appropriate?
   What is the study design?
   What is the protocol design?
   What data will be collected?

3. Group presentations and critique

Possible concept

Remote intergenerational family members
    Peripheral awareness of each other

Technology:
    Interconnected tablets
    Living room to living room
    Shareable images
    Sticky note note messages

Design concept

1-2 sentence description

Prototype state

Initial idea? Wizard-of-Oz? Interactive?

Target users

Who are they? Do you have access to them?

User activities

Real-world tasks, scenarios or experiences

Group work

What makes a good CSO study?

Study design evaluation

Does it qualify as a CSO study?

Design Concept Basis	1	builds on design concepts influenced by formative research, ideally conducted by the researcher but also from the literature.
Role of Comparison	2	ensures that each participant experiences at least two design variants in the study, e.g. different novel designs, variants within a novel design, or a baseline.
	3	structures participant activities so they can experience and compare design variants, e.g. perform equivalent tasks with each design variant.
	4	structures comparisons according to experimental design or quasi-experimental design principles, e.g. counter-balance tasks for order.
Type of Data Collected	5	records participants' comparisons and reflections on the qualitative differences in their experiences with the design variants, e.g., through interview questions.
	6	records participants' interactions with each design variant, e.g. through video recordings or high-quality cinematic logs.
	7	records quantitative data only if it helps add context to qualitative data; e.g. percentage of time participants spent in an activity.

Is it a good CSO study?

reports on substantive formative research, e.g. a well-run participatory design workshop that includes reflection by both participants and researchers.
chooses design variants that meaningfully advance the design concept(s) and avoids straw-man comparisons.
chooses and structures meaningful activities for participants, e.g., ecologically valid tasks.
justifies the protocol relative to the setting (lab or field) and comparisons being made, according to best experimental or quasi-experimental design practices.
includes well-designed interviews or surveys that elicit detailed, thoughtful comparisons by participants after exposure to the design variants.
collects rich, in-situ observational data or the best-possible alternative, e.g. remote video or substantive experience samples.
records quantitative data, if relevant, to contextualize qualitative data, e.g. participants' interactions with design elements that clarify their experiences.

Study design evaluation

Does it qualify as a CSO study?

Type of Data Analysis	8	analyzes participants' comparisons and reflections about the design variants; e.g. with reflexive thematic analysis.
	9	analyzes researchers' independent assessment of the participants' experiences; e.g. with reflexive thematic analysis.
	10	treats qualitative analysis as primary.
	11	treats quantitative analysis as secondary.
Results	12	reports findings and analysis to advance one or more design concept(s).

Is it a good CSO study?

demonstrates that participants have compared and reflected deeply about their experiences with the design variants.
leverages rich, qualitative data so that researchers can independently assess participants' reflections.
conducts and reports a rigorous qualitative analysis according to the best practices of a well-established qualitative method.
analyzes quantitative data according to the best practices of well-established quantitative methods, either or both descriptive or inferential statistics.
explicitly discusses the implications for design and how the design concept(s) should evolve, based on the study results.

Study design evaluation

Required CSO characteristics	Optional CSO characteristics	Incorrect CSO characteristics
Researchers explicitly define comparable design variants.	Lab and field settings are both appropriate.	Compares design to status quo only, outside the context of the study.
Researchers derive participant tasks based on formative research.	Participants use a talk aloud protocol.	Runs an open-ended field test with no comparisons.
Participant exposure to tasks and design variants meets experimental design best practices.	Quantitative data informs qualitative data.	Omits comparable tasks or experiences.
Participants experience at least two design variants within the study.	Researchers generate testable hypotheses.	Omits participant reflection on comparable experiences.
Participants compare and reflect upon their experiences with each design variant.		Omits researcher observation and reflection on participants' experiences.
Researchers observe participants' experiences directly or through other rich data sources.		Focuses only on performance metrics.
Researchers emphasize gathering qualitative data.		
Researchers conduct post-hoc interviews.		

Wendy E. Mackay
Inria & Université Paris-Saclay

Psychology ⟶ HCI

Joanna McGrenere
University of British Columbia

HCI ⟵ Computer Science

Useful references

Strauss & Corbin: Sociologists

"Grounded Theory" method (not theory) as a rigorous approach for analyzing observational data

Braun and Clark: Psychologists

Codified part of 'grounded theory'
Offers systematic approach for analyzing interview and observation data

Holtzblatt & Byer: Anthropologists

Contextual Inquiry focuses on design for corporate clients

Mackay & McGrenere: HCI researchers

Comparative Structured observation gathers grounded insights from participants as they compare design variants

Who is CSO for?

Greenberg & Buxton (2008)*
Usability Evaluation Considered Harmful (Some of the Time)

HCI Researchers

… also UX Designers

You have developed a new design concept
Ideally based on preliminary user studies and relevant research literature

You want to know:
Is this the "right design"? - not -
*Is the "design right"?**

Relevant skills:
Introductory experiment protocol design
Observation and interviewing techniques
Qualitative analysis, e.g. Thematic Analysis

When is CSO useful?

Design process

Mid- to late-phase of a design project

Early phase? No
- First, read literature, observe users and explore design alternatives

Middle phase? Yes
- Design and develop a design concept
  - Intuitions about use
  - No testable hypotheses

End phase? No, but …
- Not a summative evaluation method
- But offers directions for future research

What does CSO require?

Design process

Required elements

Ecologically valid task scenarios
Based on user studies and/or literature

Testable prototype
From “Wizard-of-Oz” to high-fidelity

Relevant comparison prototypes
Either “status quo” or multiple variants

Access to target users
In real or simulated environment