

Evaluation of Advanced Multimodal Interfaces

Towards New Methodologies

Camilla Forsell & Niklas Rönnerberg, June 9, 2026

Tutor instructors Linköping University, Sweden

Camilla Forsell

camilla.forsell@liu.se

User centered evaluation

Interaction design

Information visualization

Niklas Rönnerberg

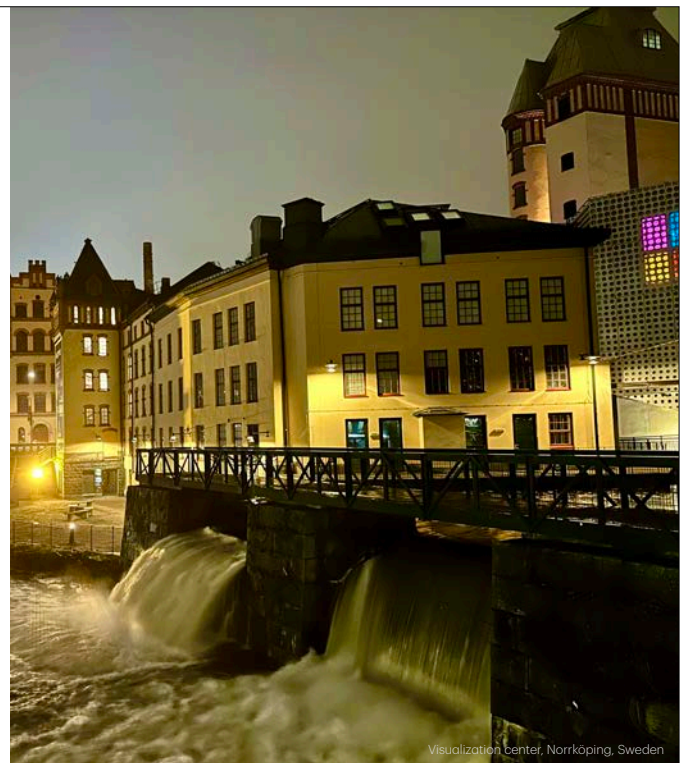
niklas.ronnerberg@liu.se

Sonification

Sonic interaction design

User evaluation

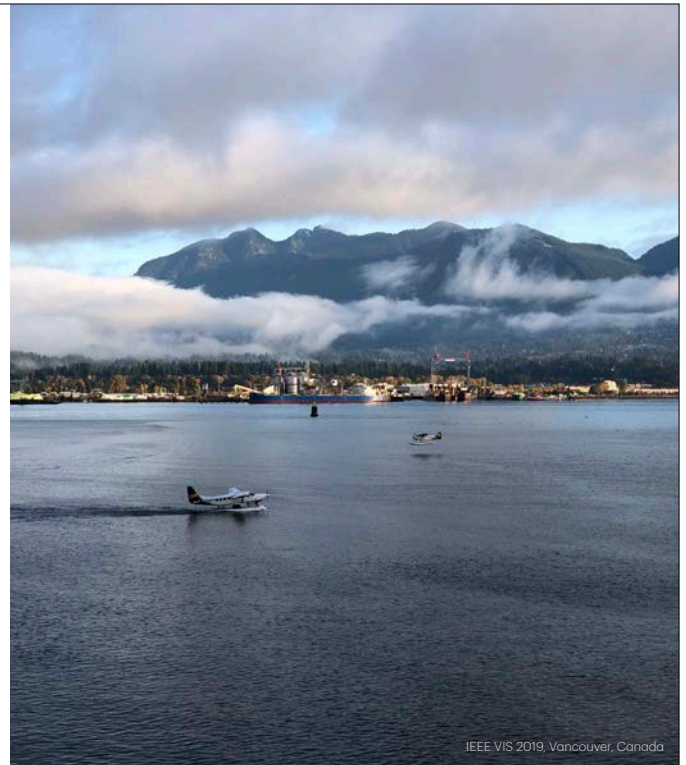
Information visualization



The workshop

Comes from tutorials...

- Eurographics 2011, Llandudno, UK
- IEEE VIS 2013, Atlanta, USA
- IEEE VIS 2016, Baltimore, USA
- IEEE VIS 2019, Vancouver, Canada
- IEEE VIS 2021, online
- IEEE VIS 2025, Vienna, Austria

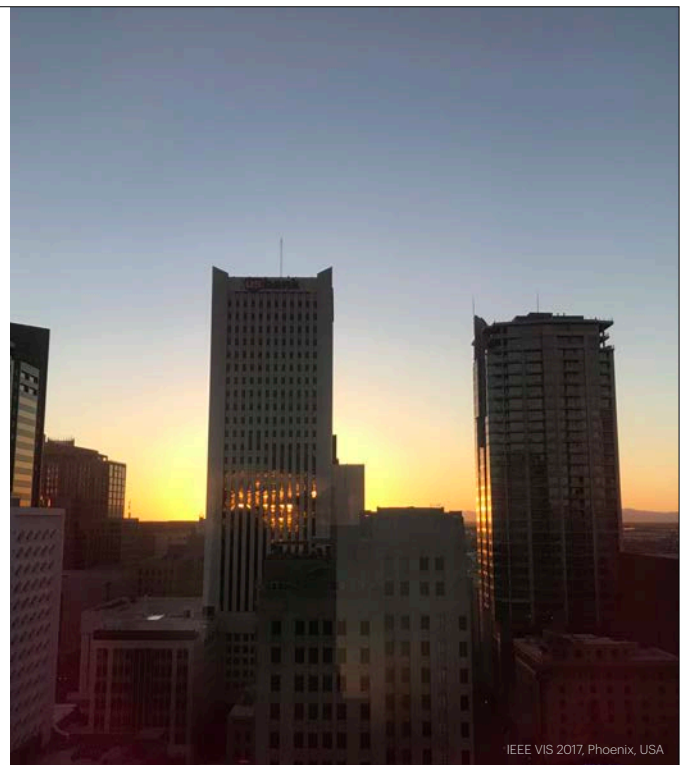


IEEE VIS 2019, Vancouver, Canada

The workshop

Comes from publications...

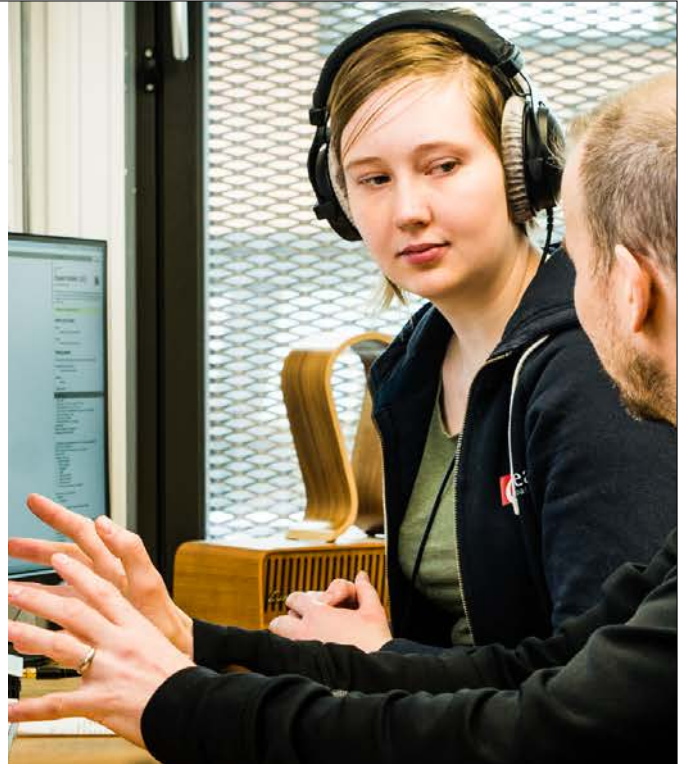
- Forsell, C., & Cooper, M. (2012, October). Questionnaires for evaluation in information visualization. In *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors- Novel Evaluation Methods for Visualization* (pp. 1-3).
- Rönnerberg, N., & Forsell, C. (2022, May 16). Questionnaires assessing usability of audio-visual representations. *AVI 2022 Workshop on Audio-Visual Analytics (WAVA22)*, Frascati, Rome, Italy. <https://doi.org/10.5281/zenodo.6555676>



IEEE VIS 2017, Phoenix, USA

Why evaluate?

- Because opinion is not evidence
- Because even plural of opinion is not evidence
- Even if everyone loves your new method
 - You don't know how, why, or what...
- So you won't know how to make it even better
- And you won't know how to use the good parts for another visualization idea
- And no-one else can use your knowledge



Why is it not standard practice?

- Evaluation is hard to do well:
 - Good study design
 - Good data selection
 - Good experimental procedure
 - Good data analysis
 - Good reporting
- Evaluation (can) take a long time
 - To evaluate something that took a day, a week or a month to implement might take a year or more to design and perform a user evaluation for...

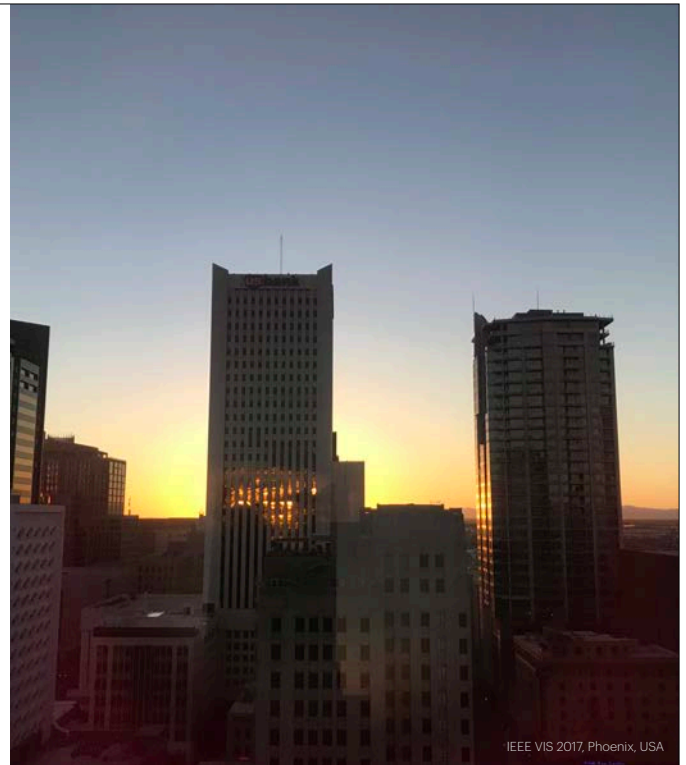


IEEE VIS 2016, Baltimore, USA

Evaluation

in multimodal interfaces...

- Is even harder than in evaluation (only) in visualisation or sonification or in user-interfaces
- Multimodal interfaces and data representations are:
 - Multi-sensory
 - Can be complementary or redundant
 - Have no standardised approaches for evaluation
 - Requires more knowledge of the researchers performing the evaluation



Agenda

What will we do today?

Introduction

Brief introductions of all attendees

User centred evaluation

Overview of current methods and challenges

Heuristics and questionnaires

Review/presentation of (some) existing evaluation instruments

Short reflections of your experiences

Group work 1

Discussion and analysis of user evaluation practices and existing questionnaires

Reporting and discussion 1

Groups share insights, followed by general discussion

Individual or group activity

Sketch and structure key ideas using affinity diagramming to identify gaps and patterns. Report in Google Doc

Group work 2

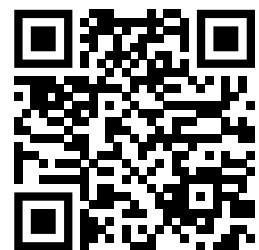
Collaborative discussion/creation of new heuristics, questions, and statements

Reporting and discussion 2

Groups share insights, followed by general discussion

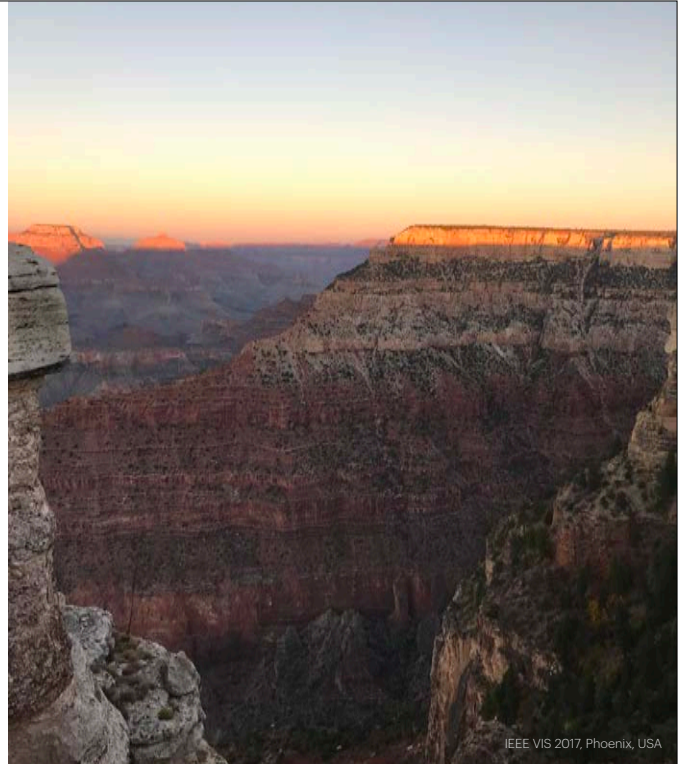
Formation of new questionnaire?

Planning next step



So,
who are you?

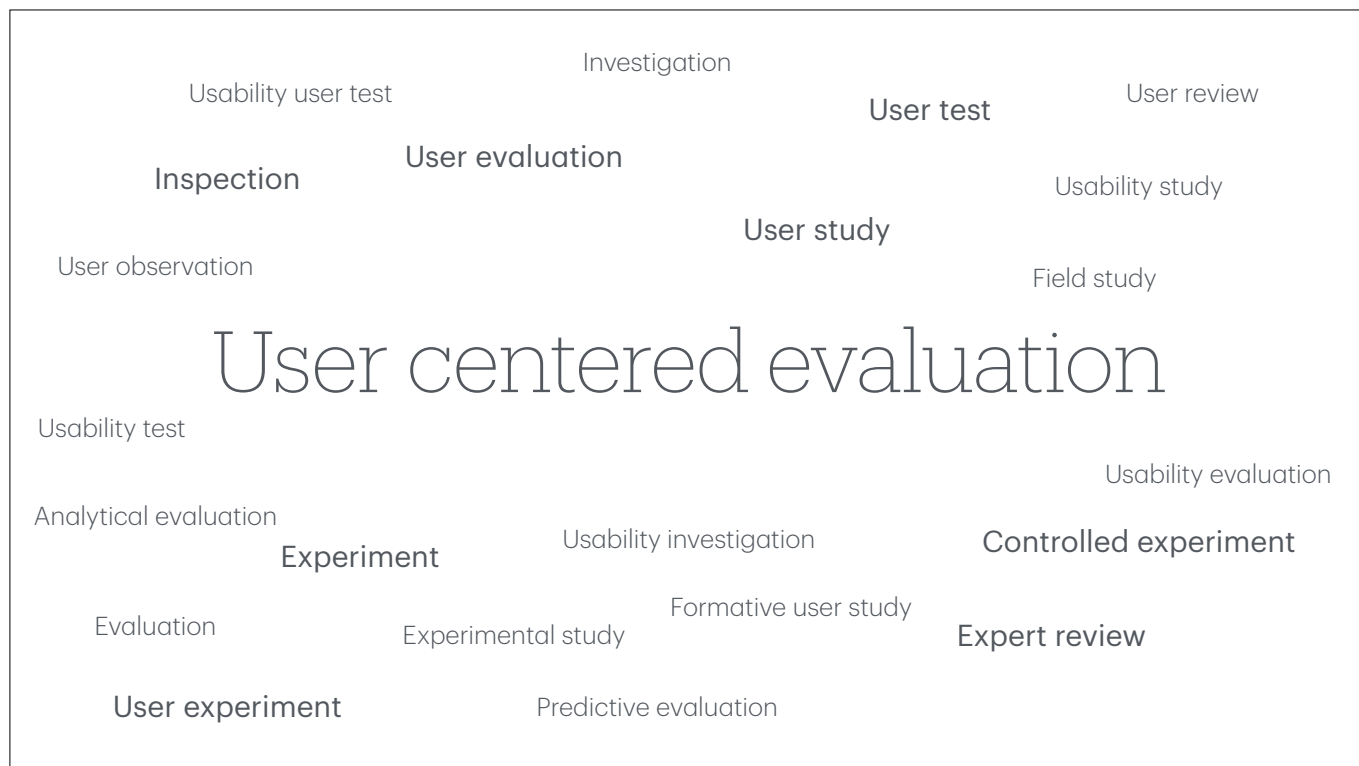
- Present briefly
 - Name
 - Affiliation
 - Main research interest



User-centered evaluation

A brief introduction

Camilla Forsell & Niklas Rönnerberg, June 9, 2026



Choosing method

Type

- Explorative
- Formative
- Summative
- Longitudinal
- Empirical
- Inspection
- Qualitative
- Quantitative
- Controlled (more so or less so)

Methods

- Observation
- Interview
- Think aloud
- Heuristic evaluation
- Measurements (performance related)
- **Questionnaire**

Methods for User-centered evaluation

Questionnaire

- A set of written statements or questions designed to collect subjective feedback
- Pre-test questionnaires collect information on participants: background, experience, work practice etc (demographic information)
- Post-test questionnaires collect reactions and attitudes, give feedback on the evaluation session
- There are two types of statements or questions:
 - 'open', where the respondent is free to answer as they wish
 - 'closed', where the respondent is asked to select from available responses often the respondent has to rate each item on a response scale (e.g. Likert scale)

Methods

Questionnaire - Likert scale

Statement: "I experienced line graph as the easiest for understanding the data"



Question: "For understanding the data, did you experience the line graph as the"



Methods

Questionnaire - Likert scale

- Be cautious with the design, and the analysis of the results (is 5 positive or negative?)
- “I can easily understand the data relationships in the line graph” (positive)
1. Strongly disagree 2. Disagree 3. Undecided 4. Agree 5. Strongly agree
- “I find it difficult to use parallel coordinates” (negative)
1. Strongly disagree 2. Disagree 3. Undecided 4. Agree 5. Strongly agree
- 5 or 7 scale?
- 1. Strongly disagree 2. Disagree 3. Agree 4. Strongly agree (4AFC)

Methods, questionnaires

Standardized questionnaires for VIS

- Questionnaires for visualization
BeauVis (aesthetic pleasure), PreVIS (readability), and Ice-T (time, insight, essence, confidence)
- Heuristics for VR/AR/MR

Methods, questionnaires

BeauVIS

- Questionnaire/rating scale to assess aesthetic pleasure

The image shows a questionnaire titled "To what extent do you agree that this visual representation is ... ?". It contains five rows, each representing a different aesthetic dimension: "enjoyable", "likable", "pleasing", "nice", and "appealing". Each row has a horizontal line with seven circular markers. The first marker is labeled "strongly disagree" and the seventh is labeled "strongly agree". A dark blue dot indicates the selected rating for each item. For "enjoyable", the dot is at the 5th marker. For "likable", it's at the 6th marker. For "pleasing", it's at the 5th marker. For "nice", it's at the 6th marker. For "appealing", it's at the 5th marker.

<https://www.aviz.fr/Research/BeauVis-Scale>

T. He., Isenberg, P., R., Dachelt and Isenberg, T. (2022). BeauVis: A Validated Scale for Measuring the Aesthetic Pleasure of Visual Representations. *IEEE Transactions on Visualization*, (29), pp 363-373.

Methods, questionnaires

PREVis

- Questionnaire/rating scale to assess perceived readability
- 4 individual scales each measuring a particular dimension of perceived readability (11 questions in total)
- Understand: the intelligibility of the encodings of the reader
- Layout: the visual clarity of the layout
- DataRead: How easily people feel they can read data values
- DataFeat: How easily people feel they can read data patterns

Cabouat, A. F., He, T., Isenberg, P., & Isenberg, T. (2024). PREVis: Perceived readability evaluation for visualizations. *IEEE Transactions on Visualization and Computer Graphics*.

PREViS

◆ Understand subscale		◆ DataRead subscale	
obvious	It is obvious for me how to read this visualization	find	I can easily find specific elements in this visualization
represent	I can easily understand how the data is represented in this visualization	identify	I can easily identify relevant information in this visualization
understand	I can easily understand this visualization	inform	I can easily retrieve information from this visualization
◆ Layout subscale		◆ DataFeat subscale	
messy	I don't find this visualization messy	visible	I find data features (for example, a minimum, or an outlier, or a trend) visible in this visualization
crowded	I don't find this visualization crowded	clear	I can clearly see data features (for example, a minimum, or an outlier, or a trend) in this visualization
distracting	I don't find distracting parts in this visualization		

☐ Strongly disagree
 ☐ Disagree
 ☐ Slightly disagree
 ☐ Neutral
 ☐ Slightly agree
 ☐ Agree
 ☐ Strongly agree

I don't know / Not applicable (if you wish, feel free to elaborate below)

<https://www.aviz.fr/Research/PREVis>

<https://www.aviz.fr/PREVis/>

PREViS

[illegible]

Methods, questionnaires

PREVis

- Recommended when participants choose: don't know/NA.
- Can also be used for more information than "just a figure/number".

PARTICIPANT ID:

PREVis questionnaire 
page 2 / 7

VISUALIZATION ID:

I don't know / Not applicable: feel free to elaborate below

It is obvious for me how to read this visualization	
I can easily understand how the data is represented in this visualization	
I can easily understand this visualization	
I don't find this visualization messy	
I don't find this visualization crowded	
I don't find distracting parts in this visualization	
I find data features (for example, a minimum, or an outlier, or a trend) visible in this visualization	
I can clearly see data features (for example, a minimum, or an outlier, or a trend) in this visualization	
I can easily find specific elements in this visualization	
I can easily identify relevant information in this visualization	
I can easily retrieve information from this visualization	

Methods, questionnaires

Standardized questionnaires for Auditory display

- Questionnaires for visualization
BeauVis (aesthetic pleasure), PreVIS (readability), and Ice-T (time, insight, essence, confidence)
- Questionnaires for auditory display
Auditory Interface User Experience test (BUZZ)
- Heuristics for VR/AR/MR

Inspection methods

Heuristic evaluation

- Inspection
 - generic term for a set of methods based on having evaluators inspect (examine and analyse) aspects of a system
- Heuristic evaluation
- Cognitive walkthrough

Nielsen, J. and Molich, R. 1990. Heuristic evaluation of user interfaces. In *Proc. ACM CHI'90 Conf*, 249–256.

Tory, M. and Möller, T. 2005. Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications* 25(5), 8-11

Inspection methods

Heuristic evaluation

- Heuristic evaluation (original form)
 - usability experts judge the compliance of the interface with recognized usability principles (known as heuristics)
 - identifies and predicts usability problems that users can be expected to encounter
 - free exploration, no specific tasks
 - around 5 evaluators, results (and severity ratings) compiled in the end

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proc. ACM CHI'90 Conf.* (s. 249–256).

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R.L. Mack (Red.), *Usability Inspection Methods* (s. 269–304). John Wiley & Sons, New York.

Inspection methods

10 Usability Heuristics for User Interface Design

- Visibility of System Status
- Match Between the System and the Real World
- User Control and Freedom
- Consistency and Standards
- Error Prevention
- Recognition Rather than Recall
- Flexibility and Efficiency of Use
- Aesthetic and Minimalist Design
- Help Users Recognize, Diagnose, and Recover from Errors
- Help and Documentation

<https://www.nngroup.com/articles/ten-usability-heuristics/>

Inspection methods

Heuristic evaluation

- Updates - in to line with new needs includes:
 - double experts (usability, visualization, domain)
 - users as evaluators/inspectors
 - specific tasks/scenarios
 - especially important if the evaluators are not end users with their own goals
 - new heuristics
 - detailed descriptions

Methods

Value-driven evaluation

- Framework for assessing the value of visualizations
- Heuristic evaluation that quantifies the value of a visualization
 - goes beyond low-level task completion tasks
 - more about understanding of data
- Complement to traditional methods with a higher-level (value-driven) focus

<https://visvalue.github.io>

Stasko, J. (2014) Value-driven evaluation of visualizations. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pp. 46-53, 2014.

Wall, E., Agnihotri, M., Matzen, L., Divis, K., Haass, M., Endert, A., & Stasko, J. (2018). A heuristic approach to value-driven evaluation of visualizations. *IEEE transactions on visualization and computer graphics*, 25(1), 491-500.

Value-driven evaluation

ICE-T

- Three hierarchical levels:
 - top level components (T + I + E + C)
 - mid-level guidelines that capture the core concepts
 - low-level heuristics that are actionable, rateable statements that reflect how a visualization achieves the guideline. They capture the core concepts of the guidelines and components

Stasko, J. (2014) Value-driven evaluation of visualizations. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pp. 46-53, 2014.

Value-driven evaluation

Low-level Heuristics – rateable statements

ICE-T

			How would you rate your agreement with the following statements?							
			Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree	N/A
Insight	The visualization exposes individual data cases and their attributes	The visualization facilitates answering questions about the data	☐	☐	☐	☐	☐	☐	☐	☐
		The visualization facilitates perceiving relationships in the data like patterns & distributions of the variables	☐	☐	☐	☐	☐	☐	☐	☐
	The visualization promotes exploring relationships between individual data cases as well as different groupings of data cases	The visualization helps generate data-driven questions	☐	☐	☐	☐	☐	☐	☐	☐
		The visualization helps identify unusual or unexpected, yet valid, data characteristics or values	☐	☐	☐	☐	☐	☐	☐	☐
	The visualization provides a new or better understanding of the data	The visualization provides useful interactive capabilities to help investigate the data in multiple ways	☐	☐	☐	☐	☐	☐	☐	☐
		The visualization shows multiple perspectives about the data	☐	☐	☐	☐	☐	☐	☐	☐
	The visualization provides opportunities for serendipitous discoveries	The visualization uses an effective representation of the data that shows related and partially related data cases	☐	☐	☐	☐	☐	☐	☐	☐
Time	The visualization affords rapid parallel comprehension for efficient browsing	The visualization provides a meaningful spatial organization of the data	☐	☐	☐	☐	☐	☐	☐	☐
		The visualization shows key characteristics of the data at a glance	☐	☐	☐	☐	☐	☐	☐	☐
	The interface supports using different attributes of the data to reorganize the visualization's appearance	The visualization supports smooth transitions between different levels of detail in viewing the data	☐	☐	☐	☐	☐	☐	☐	☐
		The visualization avoids complex commands and textual queries by providing direct interaction with the data representation	☐	☐	☐	☐	☐	☐	☐	☐
	The visualization provides mechanisms for quickly seeking specific information		☐	☐	☐	☐	☐	☐	☐	☐
Essence	The visualization provides a comprehensive and accessible overview of the data	The visualization presents the data by providing a meaningful visual schema	☐	☐	☐	☐	☐	☐	☐	☐
		The visualization facilitates generalizations and extrapolations of patterns and conclusions	☐	☐	☐	☐	☐	☐	☐	☐
	The visualization provides an understanding of the data beyond individual data cases	The visualization helps understand how variables relate in order to accomplish different analytic tasks	☐	☐	☐	☐	☐	☐	☐	☐
Confidence	The visualization helps avoid making incorrect inferences	The visualization uses meaningful and accurate visual encodings to represent the data	☐	☐	☐	☐	☐	☐	☐	☐
		The visualization avoids using misleading representations	☐	☐	☐	☐	☐	☐	☐	☐
	The visualization facilitates learning more broadly about the domain of the data	The visualization promotes understanding data domain characteristics beyond the individual data cases and attributes	☐	☐	☐	☐	☐	☐	☐	☐
		If there were data issues like unexpected, duplicate, missing, or invalid data, the visualization would highlight these issues	☐	☐	☐	☐	☐	☐	☐	☐

Inspection methods

Heuristics for metaverse user interfaces (UIs)

- A total of 52 heuristics from previous work (VR/AR/MR) were compiled together and formed a new metaverse-specific usability heuristics list containing:
 - Heuristics
 - Descriptions
 - Sub-heuristics
 - Proposed checklists



Omar, K., Fakhouri, H., Zraqou, J., & Marx Gómez, J. (2024). Usability heuristics for metaverse. *Computers*, 13(9), 222.

Inspection methods

Heuristic evaluation for virtual reality systems

- Synchronous body movements
- Physical space constraints
- Immersion
- Glitchiness
- Switch between actual and virtual world
- Cord design
- Headset comfort
- Mental comfort
- User interface design

Murtza, R., Monroe, S., & Youmans, R. J. (2017, September). Heuristic evaluation for virtual reality systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 2067-2071). Sage CA: Los Angeles, CA: SAGE Publications.

Methods, questionnaires

Standardized questionnaires for Multi-modal interfaces

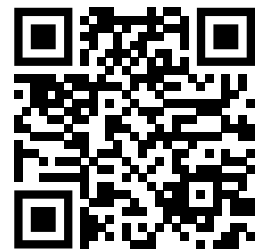
- Questionnaires for visualization
BeauVis (aesthetic pleasure), PreVIS (readability), and Ice-T (time, insight, essence, confidence)
- Questionnaires for auditory display
Auditory Interface User Experience test (BUZZ)
- Heuristics for VR/AR/MR
- Questionnaires for assessing usability
System Usability Scale (SUS), Questionnaire for User Interface Satisfaction (QUIS), Software Usability Measurement Inventory (SUMI), NASA Task Load index (NASA TLX), After-Scenario Questionnaire (ASQ), Post-Study System Usability Questionnaire (PSSUQ), Computer System Usability Questionnaire (CSUQ), Usability Metric for User Experience (UMUX and UMUX-Lite), Component model of User Experience (meCUE and meCUE2.0)
- Questionnaires for assessing immersion and engagement
Immersive Experience Questionnaire (IEQ), Gaming Engagement Questionnaire (GEQ), User Engagement Scale (UES)
- Questionnaires for aesthetics in (data representations)
Desire for Aesthetics Scale (DFAS), Centrality of Visual Product Aesthetics (CVPA), Aesthetic Emotions Scale (AESTHEMOS), Aesthetic Experience Questionnaire (AEQ)

Short reflections from you

Marco Mores

Daniel Bermúdez-Chinea

The rest of you :-)



Group work 1

Discussion and analysis of user evaluation practices
and existing questionnaires

About 30 minutes

Reporting

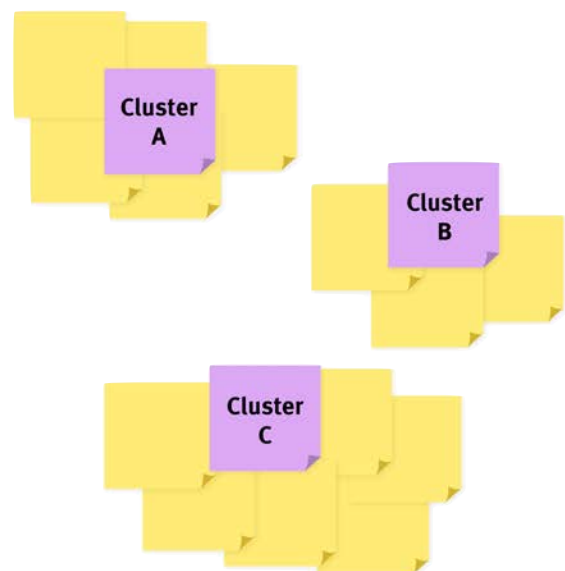
Reporting from the discussion and analysis of user evaluation practices and existing questionnaires

About 15 minutes

Affinity diagram

Finding insights...

- Systematically organize and analyze material from exercises such as brainstorming and brainwriting.
- Also for data from interviews, the 635 method, and function-driven divergence, first transferring the material to, Post-it notes.
- Brings order to collected data and provides insights.
- Gathers all ideas on a subject, and groups (categorizes/themes) ideas that are related. It also aims to remove duplicates, clarify ideas, and provide different perspectives.



<https://www.nngroup.com/articles/affinity-diagram/>

Affinity diagram

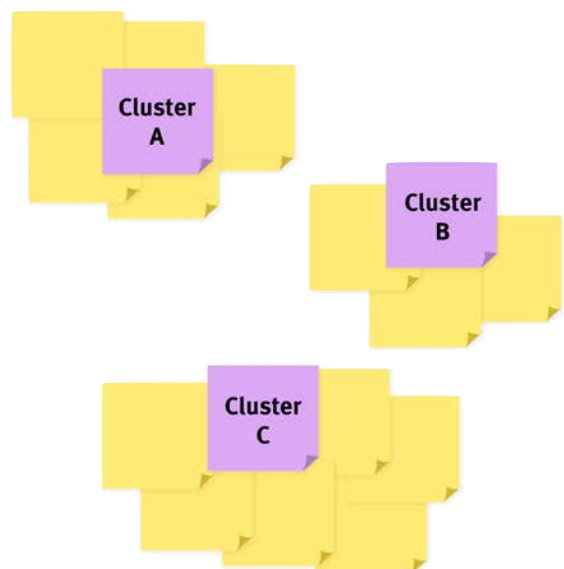
A typical affinity diagram can be created according to the following steps:

- **Step 1** Write down all ideas on notes (generated during brainstorming, analysis of interview data, etc.).
- **Step 2** Group the ideas into meaningful groups/categories. Place all ideas that are similar or related in some way into the same group.
- If several people are working together/in a group, one option is not to talk to each other at first to avoid influencing one another. A note that has already been moved (by someone) can be moved again (by someone else). Finish when no one wants to move any more notes.
- Some ideas may not relate to any others; in that case, they can stand alone.
- **Step 3** Discuss what connects the notes — what their underlying meaning or overarching theme is. Move notes if necessary.
- **Step 4** Based on Step 3, name the categories/groups. Write down a short phrase on a separate note and place it as a “Heading.” Sometimes subgroups are created by dividing groups into smaller sections.
- Themes may run through several categories (linking categories together in an overarching structure).
- **Step 5** Lines can be drawn between themes, categories, and subcategories, creating a hierarchical diagram that describes the relationships between the ideas.

Affinity diagram

Work individually or in small groups

About 30 minutes





Group work 2

Collaborative discussion/creation of new heuristics, questions, and statements

About 30 minutes

Reporting

Reporting from the discussion/creation of new heuristics, questions, and statements

About 15 minutes

Formation of new questionnaire?

Which heuristics?

Which questions/statements?

What about modularity and scope?

Camilla Forsell & Niklas Rönnerberg, June 9, 2026

Next step(s)

Finalize the multi-modal evaluation questionnaire

Use the new questionnaire

Later validation of questionnaire

Writing a paper

Who would like to be involved?

Camilla Forsell & Niklas Rönnerberg, June 9, 2026