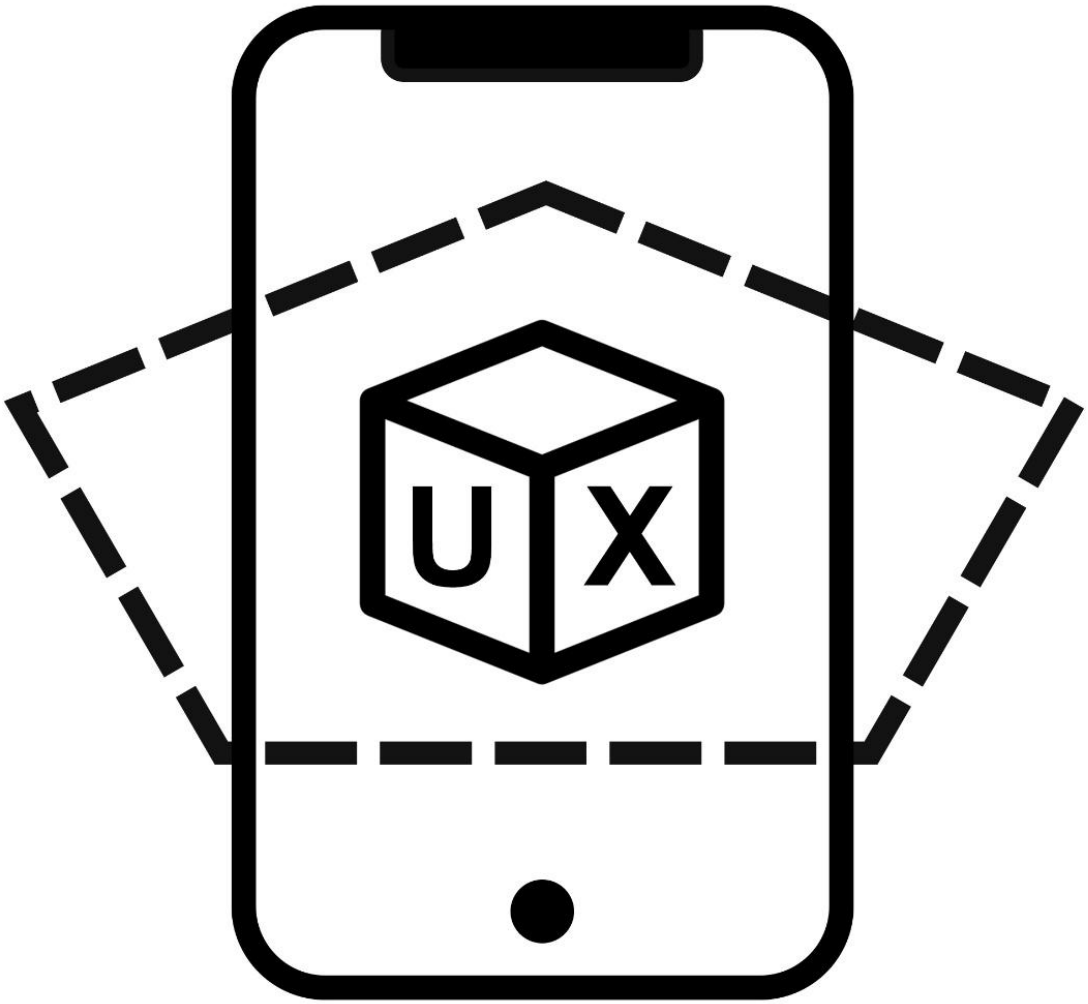


Innovative UX Evaluation Approach Handbook: Applying the UXARcis Questionnaire

© Stefan Graser

stefan.graser@ux-researchgroup.com

https://uxarcis.ueq-research.org

Version: January 2026

Content

Introduction

This handbook provides structured guidance for applying the proposed innovative UX evaluation approach in relation to the UXARcis questionnaire, the UXARclass taxonomy, and the UXARnalyze data analysis tool. The three artifacts are based on various scientific publications, including various empirical studies. Therefore, the handbook serves as a reference for providing meaningful, evidence-based suggestions for improving system design based on quantitative data.

The primary goal of the document is to support researchers and practitioners in correctly analyzing and interpreting the proposed evaluation results of this approach by using the UXARnalyze data analysis tool. The UXARnalyze, therefore, serves as a support tool for advanced data analysis, assisting users in applying the underlying evaluation methodology, enabling consistent interpretation of results, and deriving evidence-based insights.

The UXARnalyze can be accessed and downloaded on the Website or GitHub:

https://uxarcis.ueq-research.org

https://github.com/stefangraser/UXARnalyze

The document is structured in three main parts. First, the UXARcis questionnaire is introduced. Second, the UXARclass taxonomy is presented. Both provide the primary foundation of the innovative UX evaluation approach. Third, the UXARnalyze is introduced, followed by an explanation of how to use it. Moreover, some common questions and respective answers are shown at the end.

Foundation and Structure of the UXARcis Questionnaire

The UXARcis questionnaire is based on the following publications:

- Graser, S., Schrepp, M., Böhm, S. (2024). UXAR-CT – An Approach for Measuring UX for Mobile Augmented Reality Applications in Corporate Training. In: Wei, J., Margetis, G. (eds) Human-Centered Design, Operation and Evaluation of Mobile Communications. HCII 2024. Lecture Notes in Computer Science, vol 14737. Springer, Cham. https://doi.org/10.1007/978-3-031-60458-4_15
- Graser, S., Schrepp, M., Böhm, S. (2024). Construction of the UXAR-CT – a User eXperience Questionnaire for Augmented Reality in Corporate Training, CENTRIC 2024: The Seventeenth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services, Venice, Italy, ThinkMind, pp. 14-22. https://www.thinkmind.org/library/CENTRIC/CENTRIC_2024/centric_2024_1_20_30009.html
- Graser, S., Böhm, S. (2025). The Influence of System Characteristics on the User Experience of Augmented Reality Applications: An Experimental Approach in Corporate Training, 2025 International Conference on Intelligent Computing and Virtual & Augmented Reality Simulations (ICVARS), Birmingham, United Kingdom, 2025, pp. 56-65, doi: 10.1109/ICVARS66454.2025.11198670
- Graser, S., Schrepp, M., Krüger, J.M., Böhm, S. (2026). Development of an Innovative UX Measurement Approach for Augmented Reality in Corporate Training. In: Schrepp, M. (eds) HCI International 2025 – Late Breaking Papers. HCII 2025. Lecture Notes in Computer Science, vol 16341. Springer, Cham. https://doi.org/10.1007/978-3-032-13083-9_11

In the context of Augmented Reality (AR), only a limited number of domain-specific UX questionnaires currently exist, each differing in scope, structure, and conceptual focus. Moreover, no questionnaire has been explicitly designed to capture the UX of AR applications in Corporate Training (CT) contexts. Based on this, we aimed to develop a domain-specific UX questionnaire tailored to AR-based training scenarios.

For conceptual clarity, this work adopts the understanding of UX as a set of semantically well-defined UX quality aspects (UX-QA) in relation to Schrepp et al. (2023). We identified five relevant UX quality aspects based on a quantitative expert study: **Perspicuity**, **Efficiency**, **Dependability**, **Usefulness**, and **Clarity**. These aspects were operationalized in an initial questionnaire, the UXAR-CT, and empirically evaluated using real-world AR training applications following established questionnaire development procedures.

Beyond the definition of UX quality aspects, the abstraction level of measurement items represents a critical design decision. While the UXAR-CT captures UX at a higher, experience-oriented level, AR applications are also characterized by distinctive system characteristics that directly shape user perception. To address this, the UXAR-CT was extended by integrating AR-specific system characteristics, thereby enabling a more comprehensive and explanatory evaluation.

These system characteristics are represented by the ARcis criteria **Contextuality**, **Spatiality**, and **Interactivity** (Krüger 2023). The criteria are derived from established AR characteristics and reframed from a user-centered perspective. Contextuality describes the perceived integration of virtual and physical elements. Spatiality captures the assignment of stable spatial properties to virtual content within the physical environment. Interactivity refers to the user's ability to manipulate and modify virtual elements through interaction with physical or virtual objects. In the following, all scales and respective definitions are presented:

- **Perspicuity (PE):** The user easily gets familiar with the product and to learn how to use it.
- **Efficiency (EF):** The user can solve their tasks without unnecessary effort. The product reacts fast.
- **Dependability (DE):** The user feels in control of the interaction. The product reacts predictably and consistently to user commands.
- **Usefulness (US):** Using the product brings advantages to the user. Using the product saves time and effort.
- **Clarity (CL):** The user interface of the product looks ordered, tidy, and clear.
- **Contextuality (Con):** Integrated presentation of virtual and physical elements.
- **Spatiality (Spa):** Assignment of unique spatial properties to virtual elements by positioning them within the physical environment.

- **Interactivity (Int):** Modification and manipulation of virtual components by interacting with physical objects.

This results in an extended questionnaire, UXARcis, for Corporate Training scenarios. It is grounded in a well-established UX model that distinguishes between product features and product character. Accordingly, UXARcis consists of two complementary dimensions. Product features, operationalized through the ARcis criteria, and product character, represented by the five UX-QA. Together, these dimensions enable the simultaneous assessment of concrete AR system characteristics and higher-level experiential outcomes. In addition, Overall Satisfaction is included as an external anchor to contextualize evaluation results. Figure 1 illustrates the structure based on Hassenzahl's (2005) common UX model.

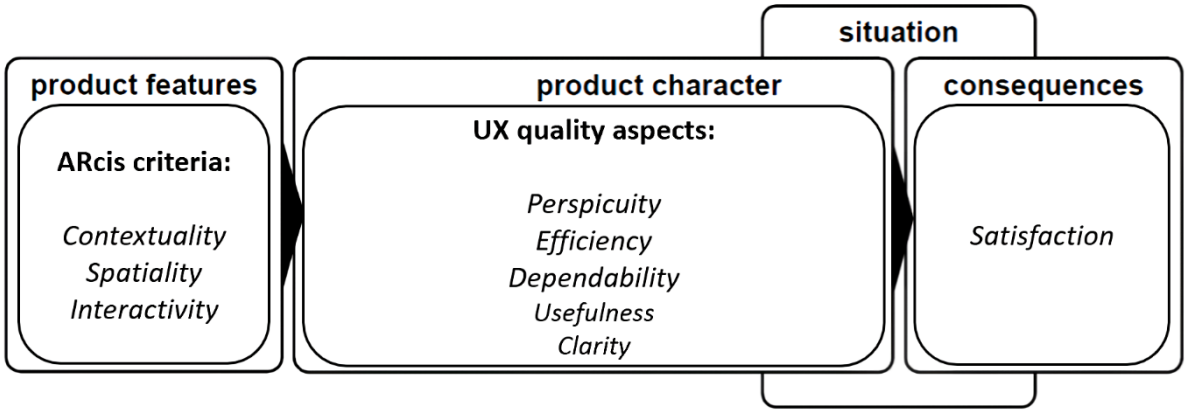


Figure 1: UXARcis model.

In total, UXARcis comprises 34 items measured on a seven-point Likert scale (From left: "Strongly disagree" = ☹" to the right "strongly agree" = ☺"). All measurement items in both English and German are provided in the Appendix.

Overall, I am satisfied with the application's support for learning.	☹	○ ○ ○ ○ ○ ○ ○	☺

The questionnaire was empirically evaluated in multiple AR training scenarios using different prototype configurations. Prior studies demonstrate a significant influence of AR-specific product features on perceived product character, providing empirical support for the underlying model. To sum up, the UXARcis offers a theoretically grounded, empirically validated UX questionnaire for AR in CT. By explicitly

integrating two levels of abstraction, the approach enables a more nuanced, interpretable, and context-sensitive assessment of user experience.

Foundation and Structure of the UXARclass Taxonomy

Quantitative evaluation results are often unspecific and provide only general indications for system design improvement, lacking concrete suggestions for improvement. Therefore, we aimed to link design recommendations with the UXARcis scales. Using a Delphi-based multimethod study, we evaluated and identified the specific relationship between the relevant design recommendations and the UXARcis scales. As a result, we provide an evidence-based connection. Tables 1 and 2 illustrate the specific classification results.

Table 1: Classification of AR design recommendation topics to the ARcis Criteria (product features)

Mean	Contextuality	Mean	Spatiality	Mean	Interactivity
5.00	16 - Keep the focus on AR experience, but use 2D-UI On-Screen elements when needed	5.00	32 - Anchored UI	4.67	5 - hand & finger gestures
4.67	1 - appropriate interplay of virtual content and physical environments	4.67	1 - appropriate interplay of virtual content and physical environments	4.67	23 - Ergonomics (avoid muscle fatigue)
4.67	21 - Text / Font	4.33	6 - Textures - Visual Realism and Appearance of Objects	4.67	9 - Handling Interruptions / Relocalization
4.67	32 - Anchored UI	4.33	10 - Surface Detection	4.67	11 - Affordance
4.33	8 - Image detection	4.33	14 - Object Manipulation	4.67	14 - Object Manipulation
4.33	10 - Surface Detection	4.00	2 - Attention directors	4.33	18 - Consider and show User's required Effort
4.33	11 - Affordance	4.00	8 - Image detection	4.33	15 - Encourage to explore
4.00	3 - Instructions	4.00	9 - Handling Interruptions / Relocalization	4.00	8 - Image detection
4.00	15 - Encourage to explore	4.00	11 - Affordance	4.00	12 - visual cues for object manipulation
3.67	12 - visual cues for object manipulation	3.33	3 - Instructions	3.67	16 - Keep the focus on AR experience, but use 2D-UI On-Screen elements when needed

3.67	2 - Attention directors	3.00	23 - Ergonomics (avoid muscle fatigue)	3.67	32 - Anchored UI
3.67	7 - Occlusion	3.00	16 - Keep the focus on AR experience, but use 2D-UI On-Screen elements when needed	3.33	1 - appropriate interplay of virtual content and physical environments
3.67	14 - Object Manipulation	3.00	21 - Text / Font	3.00	3 - Instructions
3.67	20 - Inform about Waiting Time	2.67	5 - hand & finger gestures	3.00	6 - Textures - Visual Realism and Appearance of Objects
3.33	6 - Textures - Visual Realism and Appearance of Objects	2.67	12 - visual cues for object manipulation	3.00	10 - Surface Detection
3.00	4 - Onboarding	2.67	7 - Occlusion	2.67	2 - Attention directors
3.00	9 - Handling Interruptions / Relocalization	2.67	27 - Audio Feedback	2.67	4 - Onboarding
3.00	27 - Audio Feedback	2.33	4 - Onboarding	2.67	7 - Occlusion
2.67	5 - hand & finger gestures	2.00	17 - Error prevention & recovery	2.33	21 - Text / Font
2.33	18 - Consider and show User's required Effort	2.00	18 - Consider and show User's required Effort	2.33	27 - Audio Feedback
2.33	23 - Ergonomics (avoid muscle fatigue)	1.67	20 - Inform about Waiting Time	1.67	20 - Inform about Waiting Time

Table 2: Classification of AR design recommendation topics to the UX quality aspects (product character).

Mean	Perspicuity	Mean	Efficiency	Mean	Dependability	Mean	Usefulness	Mean	Clarity
5.00	4 - Onboarding	4.67	26 - Performance	5.00	28 - Feedback	4.33	28 - Feedback	4.33	19 - Law of practice
5.00	19 - Law of practice	4.33	3 - Instructions	4.67	13 - Object Placement	4.00	11 - Affordance	4.00	2 - Attention directors

4.00	31 - headlocked content	4.33	24 - Ergonomics (avoid head & neck fatigue)	4.67	26 - Performance	3.67	9 - Handling Interruptions / Relocalization	4.00	4 - Onboarding
3.67	3 - Instructions	4.33	28 - Feedback	4.33	29 - FOV	3.67	16 - Keep the focus on AR experience, but use 2D-UI On-Screen elements when needed	4.00	7 - Occlusion
3.67	11 - Affordance	4.00	4 - Onboarding	4.00	30 - Content Placement	3.67	27 - Audio Feedback	4.00	21 - Text / Font
3.33	14 - Object Manipulation	4.00	9 - Handling Interruptions / Relocalization	3.67	4 - Onboarding	3.67	22 - Accessibility (visuals)	4.00	26 - Performance
3.33	16 - Keep the focus on AR experience, but use 2D-UI On-Screen elements when needed	4.00	9 - Handling Interruptions / Relocalization	3.67	19 - Law of practice	3.67	25 - Pause / Breaks	4.00	29 - FOV
3.33	32 - Anchored UI	3.67	2 - Attention directors	3.67	19 - Law of practice	3.33	2 - Attention directors	4.00	31 - headlocked content
3.00	2 - Attention directors	3.67	32 - Anchored UI	3.33	8 - Image detection	3.33	9 - Handling Interruptions / Relocalization	3.67	3 - Instructions
3.00	8 - Image detection	3.67	13 - Object Placement	3.33	9 - Handling Interruptions / Relocalization	3.33	30 - Content Placement	3.67	32 - Anchored UI

3.00	9 - Handling Interruptions / Relocalization	3.67	29 - FOV	3.33	20 - Inform about Waiting Time	3.33	31 - headlocked content	3.67	13 - Object Placement
3.00	15 - Encourage to explore	3.67	30 - Content Placement	3.00	7 - Occlusion	3.00	7 - Occlusion	3.33	8 - Image detection
3.00	21 - Text / Font	3.33	10 - Surface Detection	3.00	11 - Affordance	3.00	8 - Image detection	3.33	10 - Surface Detection
3.00	28 - Feedback	3.33	11 - Affordance	3.00	14 - Object Manipulation	3.00	14 - Object Manipulation	3.33	11 - Affordance
2.67	10 - Surface Detection	3.33	14 - Object Manipulation	3.00	15 - Encourage to explore	3.00	15 - Encourage to explore	3.33	22 - Accessibility (visuals)
2.67	27 - Audio Feedback	3.33	15 - Encourage to explore	3.00	27 - Audio Feedback	3.00	20 - Inform about Waiting Time	3.33	28 - Feedback
2.67	22 - Accessibility (visuals)	3.33	21 - Text / Font	3.00	32 - Anchored UI	3.00	21 - Text / Font	3.33	30 - Content Placement
2.67	29 - FOV	3.33	19 - Law of practice	3.00	24 - Ergonomics (avoid head & neck fatigue)	3.00	26 - Performance	3.00	20 - Inform about Waiting Time
2.33	20 - Inform about Waiting Time	3.33	22 - Accessibility (visuals)	2.67	2 - Attention directors	3.00	29 - FOV	3.00	24 - Ergonomics (avoid head & neck fatigue)
2.33	26 - Performance	3.33	25 - Pause / Breaks	2.67	10 - Surface Detection	2.67	3 - Instructions	2.67	14 - Object Manipulation

2.00	7 - Occlusion	3.00	16 - Keep the focus on AR experience, but use 2D-UI On-Screen elements when needed	2.67	22 - Accessibility (visuals)	2.67	4 - Onboarding	2.67	15 - Encourage to explore
2.00	25 - Pause / Breaks	3.00	20 - Inform about Waiting Time	2.33	16 - Keep the focus on AR experience, but use 2D-UI On-Screen elements when needed	2.67	10 - Surface Detection	2.67	27 - Audio Feedback
1.67	24 - Ergonomics (avoid head & neck fatigue)	3.00	7 - Occlusion	2.33	21 - Text / Font	2.67	19 - Law of practice	2.00	25 - Pause / Breaks

As a result, we can combine quantitative UX evaluation data with design practice. The user can trace UX deficits to specific design recommendations. In turn, the user also knows how specific design actions influence the resulting UX.

UXARnalyze Data Analysis Tool

The data analysis tool aims to provide an easy-to-use, practitioner-oriented solution for UX-AR researchers to automate the analysis of UXARcis questionnaire data based on the UXARclass taxonomy. In particular, the tool not only provides descriptive data operations, such as mean scores and correlations, but also provides evidence-based design feedback by connecting the UX scales to the relevant AR design recommendations based on the UXARclass taxonomy. Moreover, the calculation of an initial benchmark based on cut points is included.

We aimed to provide a freely accessible and usable tool, unlike most others. Therefore, the focus during development was not on a scalable application, but purely on practical utility. We followed the structure of the UEQ data analysis tool, relying on an Excel-based workflow ***(Microsoft Office 2024)***. The UXARnalyze was evaluated based

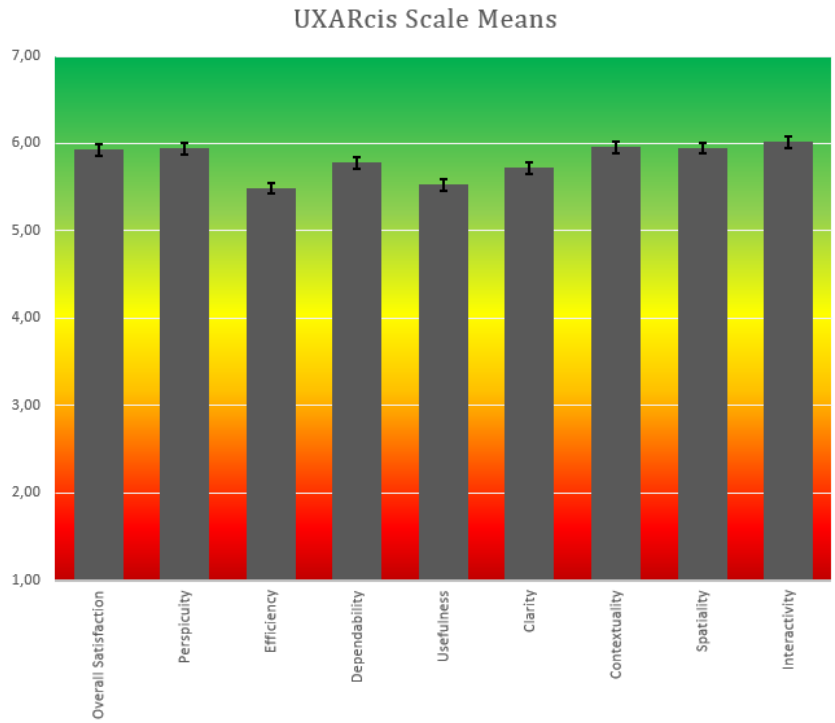
on a qualitative expert study. The detailed structure and data operations are described below.

Introduction & Data Input

The first worksheet provides an introduction to the tool. The user has to specify the device characteristics of the evaluated application. In particular, users can choose between handheld (HH), head-mounted (HMD), or both. Moreover, all respective worksheets are briefly described. The worksheet ***Data*** contains a table listing the measurement items in the order they appear in the questionnaire. This is the only sheet to be modified, where the raw data is to be entered.

Quantitative Statistical Results

The worksheet ***Quant_Results*** provides descriptive data analysis, including scale means with their visualizations, variances, and correlations.



Practical Improvement Suggestions

The worksheet ***Pract_Results*** compares the scale means against the initial cut-point benchmark. In the case the mean score is above the cut point and, therefore, fulfilled, the indication \textit{"No action needed"} is given. If a mean score is below the cut point and, therefore, critical, users can click the corresponding button "**Show relevant design recommendations**" to open the respective worksheet ***DesRec_scale***

abbreviation with the relevant AR design recommendations for this scale. This can be considered a type of recommender function. The respective worksheets are based on UXARclass and included in the database. We describe this later on. Thus, users are directly given an indication of possible deficits.

	Cut Point value for product character scales: *(for details, see worksheet CutPoint_productcharacter)* 5,25					*Cut Point value for product features scales:* *(for details, see worksheet CutPoint_productfeatures)* 5,25		
Scale	*Perspicuity*	*Efficiency*	*Dependability*	*Usefulness*	*Clarity*	*Contextuality*	*Spatiality*	*Interactivity*
Means	5,94	5,49	5,78	5,52	5,72	5,95	5,95	6,02
CutPoint Comparison	fulfilled	fulfilled	fulfilled	fulfilled	fulfilled	fulfilled	fulfilled	fulfilled
	No action needed	No action needed	No action needed	No action needed	No action needed	No action needed	No action needed	No action needed
Click to see practical suggestions -->								

Further Data Analysis

Moreover, we included further data analysis that may be helpful for interpretation. In the sheet ***Scale_Consistency***, we analyzed scale reliability using the two common coefficients, *Cronbach's Alpha* and *Guttmann's Lambda2* . We further show the specific item correlations in the worksheet with the same name. Moreover, we show all confidence intervals for scale means and single items. We also show the distribution of answers for each item. To detect suspicious data, we compared all items of a UX scale. The idea is to check how much the best and worst evaluations of an item in a specific scale differ. If there is a big difference, answers may be seen as critical and therefore need to be deleted. Lastly, we present a sheet for sample size estimation. For this, we estimate the required sample size by using the current sample's standard deviation to approximate the participant variability and then compute, for a chosen precision (E) and error probability (P), how many participants are needed so that the estimated scale mean lies within the desired confidence interval with probability 1–P.

Initial Benchmark based on CutPoints

The benchmark consists of two worksheets, *CutPoint_productcharacter* and *CutPoint_productfeatures*, containing the calculation of the cut points for the two UX dimensions.

Database

The database contains a worksheet with the whole structure of the UXARcis and the final version of the questionnaire with randomized item order. Moreover, we have included all 284 relevant AR design recommendations in the worksheet ***DesignRec***. Moreover, the worksheet ***DT*** shows the ordered items based on the UX scales and shows the respective mean values of the scales per participant. Data operations rely on

this sheet. Lastly, we present eight worksheets with the ***DesRec_scale abbreviation*** that users can access via the button in the sheet ***Pract_Results*** at a critical value for the respective UX scale. In the sheet, the user can choose how many topic categories to display, starting with the most relevant. Each topic category includes all design recommendation topics with the same mean score. For example, choosing number one shows all relevant AR design recommendations score highest based on the UXARclass taxonomy. Users are shown the relevant topic and the corresponding design recommendations for the device characteristics selected in the first worksheet introduction. As a result, users can review all displayed design recommendations in the form of a checklist and apply them to their application to address the UX deficit. If a recommendation was checked and/or applied, it can be ticked, resulting in a green mark-up.

Usage and Interpretation

How to use the UXARnalyze tool

Using the UXARnalyze is straightforward. As the goal of the tool is to make the analysis of the UXARcis questionnaire data as easy as possible for you, you don't have to change anything within the respective worksheets. You only have to **(1) specify the evaluated device** and **(2) input your evaluation data**. Based on this, you can move from sheet to sheet to get the respective data relevant to your goals. In particular, the two main steps are described in the following:

Step (1): Specification of device:

In the first worksheet, Introduction, there is a **red field**. By clicking on the field, a drop-down selection option is visible. This is also indicated by the sign **"Choose here!"**. The following Figure shows the excerpt from the tool.

Figure tbd: Device specification.

Regarding selection options, the user can choose between **HH (handheld)**, **HMD (head-mounted)**, or **both (HH/HMD)**. Please choose and select the respective device that was evaluated in your testing.

The selection is particularly relevant as the AR design recommendations are classified based on their relevance and applicability to the device.

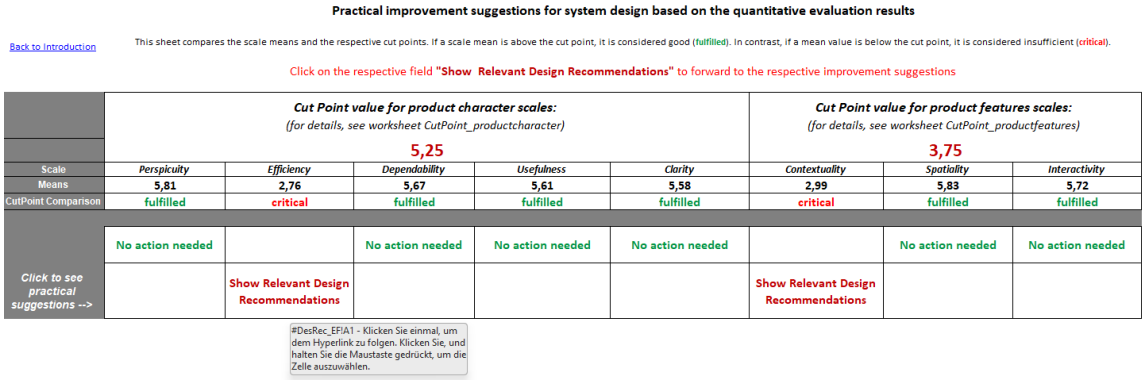
Step (2): Input the evaluation data:

In the second step, you have to input the raw evaluation data into the worksheet *Data*. For this, just **copy and paste** the data into the sheet. Please note: **This is the only sheet you need to modify. Do not change the content of the other worksheets!**

Based on this, all relevant statistics are calculated and shown within the respective worksheets.

How to interpret results

The tool contains different statistical measures. As described, the main contribution of the UXARnalyze tool goes beyond the illustration of pure quantitative data by matching the quantitative statistical results with concrete design actions. Therefore, the relevant sheet is *Pract_Results*. Within this worksheet, the calculated cut point representing the initial benchmark is compared to the scale means. If a scale mean is above the benchmark, it can be considered sufficiently positive, and no action is needed. If the scale mean is below the benchmark, indicating a bad experience, it is marked as **critical**. Based on this, the user is presented with a button "**Show relevant Design Recommendations**" linked to the subsidiary worksheets *DesRec_ScaleAbbreviation*. By clicking on the link, the user is directed to the respective sheet.



Practical improvement suggestions for system design based on the quantitative evaluation results

Back to Introduction

This sheet compares the scale means and the respective cut points. If a scale mean is above the cut point, it is considered good (fulfilled). In contrast, if a mean value is below the cut point, it is considered insufficient (critical).

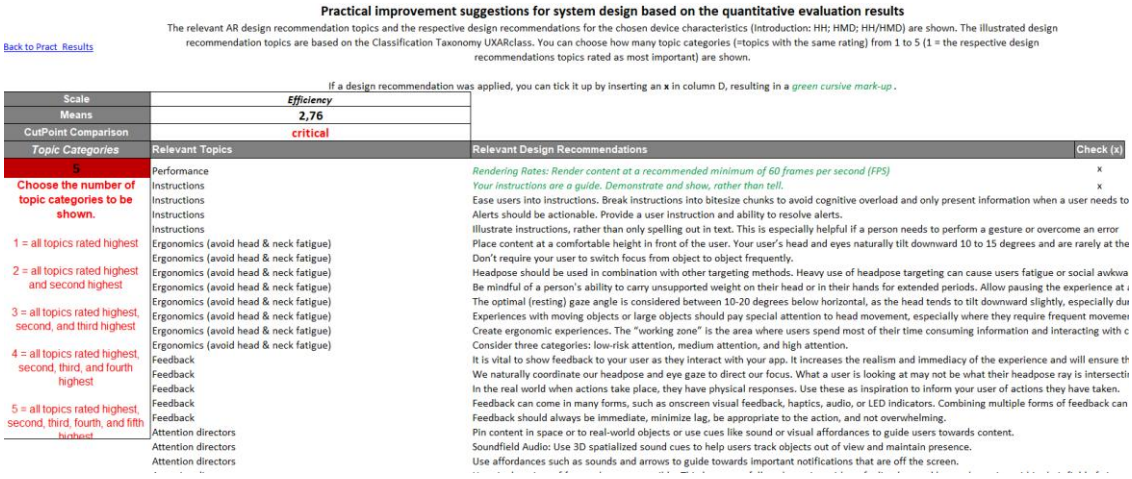
Click on the respective field **"Show Relevant Design Recommendations"** to forward to the respective improvement suggestions

	Cut Point value for product character scales: (for details, see worksheet CutPoint_productcharacter)					*Cut Point value for product features scales:* (for details, see worksheet CutPoint_productfeatures)		
	5,25					3,75		
Scale	*Perspicuity*	*Efficiency*	*Dependability*	*Usefulness*	*Clarity*	*Contextuality*	*Spatiality*	*Interactivity*
Means	5,81	2,76	5,67	5,61	5,58	2,99	5,83	5,72
CutPoint Comparison	fulfilled	critical	fulfilled	fulfilled	fulfilled	critical	fulfilled	fulfilled
	No action needed		No action needed	No action needed	No action needed		No action needed	No action needed
Click to see practical suggestions -->		Show Relevant Design Recommendations				Show Relevant Design Recommendations		

#DesRec_EF!A1 - Klicken Sie einmal, um dem Hyperlink zu folgen. Klicken Sie, und halten Sie die Maustaste gedrückt, um die Zelle auszuwählen.

Figure tbd: Excerpt from the worksheet *Pract_Results*.

Within the worksheets *DesRec_ScaleAbbreviation*, the design recommendations sorted by relevance based on the UXARclass in relation to the specified device are shown. The illustrated design recommendations can be considered similar to a heuristic checklist by the respective researchers. Regarding interpretation, the user can assume that the UX scale can be improved by checking and applying the design recommendations. If a design recommendation was applied, it can be marked by inserting an "*x*", resulting in a green mark-up.



Practical improvement suggestions for system design based on the quantitative evaluation results

Back to Pract_Results

The relevant AR design recommendation topics and the respective design recommendations for the chosen device characteristics (Introduction: HH; HMD; HH/HMD) are shown. The illustrated design recommendation topics are based on the Classification Taxonomy UXARclass. You can choose how many topic categories (=topics with the same rating) from 1 to 5 (1 = the respective design recommendations topics rated as most important) are shown.

If a design recommendation was applied, you can tick it up by inserting an **x** in column D, resulting in a *green cursive mark-up*.

Scale	*Efficiency*		
Means	2,76		
CutPoint Comparison	critical		
Topic Categories	Relevant Topics	Relevant Design Recommendations	Check (x)
5	Performance	*Rendering Rates: Render content at a recommended minimum of 60 frames per second (FPS)*	x
Choose the number of topic categories to be shown.	Instructions	*Your instructions are a guide. Demonstrate and show, rather than tell.*	x
	Instructions	Ease users into instructions. Break instructions into bitesize chunks to avoid cognitive overload and only present information when a user needs to	
	Instructions	Alerts should be actionable. Provide a user instruction and ability to resolve alerts.	
	Instructions	Illustrate instructions, rather than only spelling out in text. This is especially helpful if a person needs to perform a gesture or overcome an error	
1 = all topics rated highest	Ergonomics (avoid head & neck fatigue)	Place content at a comfortable height in front of the user. Your user's head and eyes naturally tilt downward 10 to 15 degrees and are rarely at the	
2 = all topics rated highest and second highest	Ergonomics (avoid head & neck fatigue)	Don't require your user to switch focus from object to object frequently.	
	Ergonomics (avoid head & neck fatigue)	Headpose should be used in combination with other targeting methods. Heavy use of headpose targeting can cause users fatigue or social awkwa	
	Ergonomics (avoid head & neck fatigue)	Be mindful of a person's ability to carry unsupported weight on their head or in their hands for extended periods. Allow pausing the experience at	
3 = all topics rated highest, second, and third highest	Ergonomics (avoid head & neck fatigue)	The optimal (resting) gaze angle is considered between 10-20 degrees below horizontal, as the head tends to tilt downward slightly, especially du	
	Ergonomics (avoid head & neck fatigue)	Experiences with moving objects or large objects should pay special attention to head movement, especially where they require frequent moveme	
4 = all topics rated highest, second, third, and fourth highest	Ergonomics (avoid head & neck fatigue)	Create ergonomic experiences. The "working zone" is the area where users spend most of their time consuming information and interacting with c	
	Ergonomics (avoid head & neck fatigue)	Consider three categories: low-risk attention, medium attention, and high attention.	
	Feedback	It is vital to show feedback to your user as they interact with your app. It increases the realism and immediacy of the experience and will ensure th	
	Feedback	We naturally coordinate our headpose and eye gaze to direct our focus. What a user is looking at may not be what their headpose ray is intersecti	
	Feedback	In the real world when actions take place, they have physical responses. Use these as inspiration to inform your user of actions they have taken.	
5 = all topics rated highest, second, third, fourth, and fifth highest	Feedback	Feedback can come in many forms, such as onscreen visual feedback, haptics, audio, or LED indicators. Combining multiple forms of feedback can	
	Feedback	Feedback should always be immediate, minimize lag, be appropriate to the action, and not overwhelming.	
	Attention directors	Pin content in space or to real-world objects or use cues like sound or visual affordances to guide users towards content.	
	Attention directors	Soundfield Audio: Use 3D spatialized sound cues to help users track objects out of view and maintain presence.	
	Attention directors	Use affordances such as sounds and arrows to guide towards important notifications that are off the screen.	

Figure tbd: Excerpt from the worksheet *DesRec_EF*.

Of course, it is unrealistic and impractical to check a large number of statements in practice. Therefore, the user can further select how many relevant design recommendation topics and the respective design recommendations are shown. In particular, the user can choose between 1 and 5 clusters. Each cluster contains the respective design recommendation topics that are scored similarly in the UXARclass taxonomy. Thus, number 1 refers to the topic(s) with the highest classification score.

Further Information & Typical Questions

Can I change some items?

You should **NOT** change single items or leave out some items in a scale! If you do this, it is very difficult to interpret your results. For example, the cut point values for the initial benchmark, calculated based on the original items, should not be used, simply because the answers are not comparable.

Do I need all scales?

You can leave out complete scales, i.e., delete all items of a certain scale from the questionnaire. This can make sense to shorten the questionnaire in cases where it is clear that a certain scale is not of interest.

How long do participants need to fill out the questionnaire?

Applying the UXARcis does not require much effort. Usually, 8-10 minutes are sufficient for a participant to read the instructions and to complete the questionnaire.

How much data do I need?

The more data you have collected, the better and more stable the scale means will be, and thus the more accurate the conclusions you draw from these data. However, it is not possible to give a minimum number of data points you need to collect to get reliable results. How much data you need also depends on the level of agreement among the users who participate in the questionnaire (the scale's standard deviation). The more they agree, i.e., the lower the standard deviation of the answers to the items is, the less data you need for reliable results. For typical products evaluated so far, around 20-30 people already give relatively stable results.

The Excel-based data analysis tool contains a worksheet named Sample_Size. Here, the standard deviation per scale is used to estimate how many data points you need to reach a certain precision (measured by the width of the confidence interval) in your measurement. Obviously, the precision depends on the conclusions you want to draw

from the data. For typical product evaluations, a precision of 0.5 seems to be adequate (see the detailed explanations in the worksheet).

What does the domain Corporate Training include?

Corporate Training (CT) provides a heterogeneous landscape, including multiple areas, application types, and training methods. Against this, we have further broken down CT into the different training methods. Martin et al. (2013)[1] defined 13 different methods, evaluated by Butaslac et al. (2023)[2] in the context of AR. Based on this, **programmed instruction**, presenting learning materials in a structured, step-by-step manner to learners, is one of the most widely applied methods. Moreover, previous research indicates that AR-programmed instruction applications are primarily used for training, assembly tasks, and maintenance or repair operations. Therefore, we relate CT to **AR-programmed instruction** applications for further specification and limitation.

How do we understand the term design recommendation?

We want to clarify terms and definitions in advance to establish a common understanding. Over the last few decades, different terms such as *principle*, *guideline*, or *heuristics* have been established. Principles are formulated in general terms. *Guidelines*, on the other hand, are more specific. They can be translated into heuristics for evaluating systems. Fu et al. (2016)[3] consolidated the following descriptions:

• **Principle**: A fundamental rule or law, derived inductively from extensive experience and/or empirical evidence, which provides design process guidance to increase the chance of reaching a successful solution

[1] Martin, B., Kolomitro, K., Lam, T.: Training Methods: A Review and Analysis. Human Resource Development Review 13, 11–35 (Feb 2013). https://doi.org/10.1177/1534484313497947

[2] Butaslac, I.I.M., Fujimoto, Y., Sawabe, T., Kanbara, M., Kato, H.: Systematic Review of AugmentedReality Training Systems. IEEE Transactions on Visualization and Computer Graphics 29(12),5062-5082. https://ieeexplore.ieee.org/document/9866555/?arnumber=9866555, conference Name: IEEE Transactions on Visualization and Computer Graphics

[3] Fu, K. K., Yang, M. C., and Wood, K. L. (2016). "Design Principles: Literature Review, Analysis, and Future Directions." ASME. J. Mech. Des. October 2016; 138(10): 101103. https://doi.org/10.1115/1.4034105

• **Guideline**: A context-dependent directive, based on extensive experience and/or empirical evidence, which provides design process direction to increase the chance of reaching a successful solution.

• **Heuristics**: A context-dependent directive, based on intuition, tacit knowledge, or experiential understanding, which provides design process direction to increase the chance of reaching a satisfactory but not necessarily optimal solution.

We use the term **design recommendations**, including all three terms, to establish a common understanding.

What should I do if I have further questions?

Please contact the corresponding author, Stefan Graser (stefan.graser@ux-researchgroup.com).

Appendix

Measurement Items (English)

The Content in the brackets [] **must be adjusted** in relation to the AR application.

S	Overall, I am satisfied with the application's support for learning.
PE1	It is easy/simple to learn how to use the application
PE2	The information in the application is easy to understand
PE3	The operation of the application is logical
EF1	Using the application for learning is practical
EF2	The application helps me to learn faster
EF3	The application improves my learning and work performance
DE1	The application is easy to control
DE2	I always have control over the application at every step.
DE3	It is easy to find your way around the application.
US1	The application helps me to learn
US2	It is a great advantage to use the application when learning
US3	I find the application useful for learning
CL1	The information on the display is clearly laid out
CL2	The information on the display is clear
CL3	It's easy to find the information I need
Con1	The [virtual elements] in the [real environment] were important for learning.
Con2	The [virtual elements] matched well with the [real learning situation]
Con3	The [virtual elements] were connected to [physical objects] at the appropriate time
Con4	The combination of [physical and virtual elements] was useful for learning
Con5	The simultaneous use of [physical and virtual elements] was useful for learning
Con6	The linking of [physical and virtual elements] was easy to understand
Spa1	The spatial connection between [physical and virtual elements] was easy to understand
Spa2	The spatial connection of the [virtual elements] was easy to understand
Spa3	The spatial connection between [physical and virtual elements] was useful for learning
Spa4	The spatial connection of the [virtual elements] was useful for learning
Spa5	The [virtual elements] in 3D was useful for learning
Spa6	The placement of the [virtual elements] was useful for learning
Int1	The interaction with the [virtual elements] was useful for learning
Int2	The interaction with the [physical objects] was useful in learning
Int3	The [virtual elements] were easy to use
Int4	The [physical objects] were easy to use
Int5	When interacting with the [physical object], I received useful feedback from [virtual elements]
Int6	When interacting with the [virtual element], I received useful feedback from [physical objects]

Measurement Items (German)

S	Insgesamt bin ich mit der Unterstützung der Anwendung fürs Lernen zufrieden.
PE1	Es ist leicht/einfach zu lernen, wie man die Anwendung benutzt.
PE2	Die Informationen der Anwendung sind einfach zu verstehen.
PE3	Die Bedienung der Anwendung ist logisch.
EF1	Die Nutzung der Anwendung fürs Lernen ist praktisch.
EF2	Die Anwendung hilft mir schneller zu lernen.
EF3	Die Anwendung verbessert meine Lern- und Arbeitsleistung.
DE1	Die Anwendung ist einfach zu kontrollieren.
DE2	Ich habe bei allen Schritten immer die Kontrolle über die Anwendung.
DE3	Es ist einfach sich in der Anwendung zurechtzufinden.
US1	Die Anwendung hilft mir beim Lernen.
US2	Es ist von großem Vorteil, die Anwendung beim Lernen zu nutzen.
US3	Ich finde die Anwendung fürs Lernen nützlich.
CL1	Die Darstellung der Informationen im Display ist übersichtlich.
CL2	Die Darstellung der Informationen im Display ist klar.
CL3	Es ist einfach, die Informationen zu finden, die ich benötige.
Con1	Die [virtuellen Elemente] in der [realen Umgebung] waren wichtig für das Lernen.
Con2	Die [virtuellen Elemente] passten gut zur [realen Lernsituation]
Con3	Die [virtuellen Elemente] waren zur passenden Zeit mit [physischen Objekten] verbunden
Con4	Die Verknüpfung von [physischen und virtuellen Elementen] war nützlich für das Lernen
Con5	Die gleichzeitige Verwendung von [physischen und virtuellen Elementen] war nützlich für das Lernen
Con6	Die Verknüpfung von [physischen und virtuellen Elementen] war einfach zu verstehen
Spa1	Die räumliche Verbindung zwischen [physichen und virtuellen Elementen] war einfach zu verstehen
Spa2	Die räumliche Verbindung der [virtuellen Elementen] war einfach zu verstehen
Spa3	Die räumliche Verbindung zwischen [physichen und virtuellen Elementen] war nützlich beim Lernen
Spa4	Die räumliche Verbindung der [virtuellen Elementen] war nützlich beim Lernen
Spa5	Die [virtuellen Elemente] in 3D war nützlich für das Lernen
Spa6	Die Platzierung der [virtuellen Elemente] war nützlich für das Lernen
Int1	Die Interaktion mit den [virtuellen Elementen] war nützlich beim Lernen
Int2	Die Interaktion mit den [physischen Objekten] war nützlich beim Lernen
Int3	Die [virtuellen Elementen] waren einfach zu bedienen
Int4	Die [physischen Objekte] waren einfach zu bedienen
Int5	Bei der Interaktion mit dem [physischen Objekt] habe ich nützliche Rückmeldung durch [virtuelle Elemente] erhalten
Int6	Bei der Interaktion mit dem [virtuellen Element], habe ich nützliche Rückmeldung durch [physische Objekte] erhalten

Final, randomized UXARcis questionnaire

Code	Items
S	Overall, I am satisfied with the application's support for learning.
EF3	The application improves my learning and work performance
CL2	The information on the display is clear
PE3	The operation of the application is logical
CL1	The information on the display is clearly laid out
DE1	The application is easy to control
EF1	Using the application for learning is practical
PE2	The information in the application is easy to understand
PE1	It is easy/simple to learn how to use the application
US1	The application helps me to learn
DE2	I always have control over the application at every step.
DE3	The application always responds comprehensible
CL3	It's easy to find the information I need
EF2	The application helps me to learn faster
US2	It is a great advantage to use the application when learning
US3	I find the application useful for learning
Spa5	The [virtual elements] in 3D was useful for learning
Int4	The [physical objects] were easy to use
Spa2	The spatial connection of the [virtual elements] was easy to understand
Con4	The combination of [physical and virtual elements] was useful for learning
Con2	The [virtual elements] matched well with the [real learning situation]
Spa4	The spatial connection of the [virtual elements] was useful for learning
Spa1	The spatial connection between [physical and virtual elements] was easy to understand
Con1	The [virtual elements] in the [real environment] were important for learning.
Int2	The interaction with the [physical objects] was useful in learning
Int6	When interacting with the [virtual element], I received useful feedback from [physical objects]
Con6	The linking of [physical and virtual elements] was easy to understand
Int3	The [virtual elements] were easy to use
Int1	The interaction with the [virtual elements] was useful for learning
Con5	The simultaneous use of [physical and virtual elements] was useful for learning
Spa6	The placement of the [virtual elements] was useful for learning
Spa3	The spatial connection between [physical and virtual elements] was useful for learning
Con3	The [virtual elements] were connected to [physical objects] at the appropriate time
Int5	When interacting with the [physical object], I received useful feedback from [virtual elements]