Identifying UEQ+ Scales for the categories Dashboard and VR in a quantitative study

ANONYMOUS AUTHOR(S)

This paper introduces new product categories, Dashboard and Virtual Reality, to the User Experience Questionnaire (UEQ+). The UEQ+ is a standardized questionnaire that provides a set of scales considered to be most important for evaluating the user experience of a product family. We have conducted a quantitative study to determine the most important scales. For this purpose, a procedure that had been used in previous studies to determine the most important User Experience scales for the categories Games and Learning Platforms was reproduced. These product categories were integrated to ensure that the conditions of our study were similar. The results are presented, critically discussed, and compared with previous studies. The derived procedure is assessed for its validity and presented in an open repository. As a conclusion, we present the most important scales for UEQ+ evaluations for Dashboard and Virtual Reality products.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: UEQ+, dashboard, VR

ACM Reference Format:
Anonymous Author(s). 2023. Identifying UEQ+ Scales for the categories Dashboard and VR in a quantitative study. In . ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Measuring the impact of computer-supported systems or software in HCI is a challenging task. Different disciplines are addressing this field differently in their evaluation: either the interaction of the human with the computer or the computer with the human is observed [9].

Surveys and questionnaires are the primary source[10] of observation of interaction of the human with the computer. To compare the quality of interaction between human and computer, a set of comparable interaction settings need to exist. For example, if user excitement is measured in a game-based learning scenario, then the results can seem very thrilling without a baseline of game-based learning applications in general. Standardized questionnaires are mostly used to create this baseline because of a predefined set of indicators. The interaction scenario could be addressed by measuring Human Factors[9], like the questionnaire of the Technology Acceptance Model (TAM) [4], or by specific ergonomic aspects[9], like the User Experience Questionnaire (UEQ) [7]. These and other standardized questionnaires were created and tested in the field, to compare similar interaction settings with each other. For the following article we have chosen to focus on the UEQ+, a modularizable expansion of the UEQ, to analyze ergonomic aspects of applications and software products instead of human factors and whole interaction scenarios.

Manuscript submitted to ACM

The UEQ offers a baseline in a variety of general User Experience (UX) aspects [13]. In the context of the UEQ these UX aspects are called scales: *Attractiveness* (the overall impression of the product), *pragmatic qualities* (Perspicuity - how easy is it to get familiar with the product?; Efficiency - can tasks be solved with minimum effort?; Dependability - Does the user have the feeling of controlling the interaction?) and *hedonic qualities* (Stimulation - Is it exciting to use the product?; Novelty - Is the product innovative?) [19]. In further progression, the UEQ+ was created by integrating other aspects and scales in their questionnaire, like the efficiency of use, ease of learning, controllability, error tolerance, intuitive use, visual complexity, usefulness, the fun of use, identity, aesthetics of the visual design, novelty of the product concept, or content quality [11]. Collecting data for all of these categories would result in a questionnaire exceeding a reasonable length, and not every scale is important for every product [11]. Therefore, a concept to modularize the scales for specific products was integrated into the UEQ+. Comparability can be gained by forming product categories for which relevant scales have been determined in advance. This leads to a quantitative evaluation by questionnaire which is easy to use and has a focus on user experience. The number of product categories listed in the manual for using the UEQ+[12] does not claim to be complete. One way to include a new product category is a short empirical study of the target user group to measure the importance of specific performance indicators. In order to receive a baseline for our future research activities, we have decided to create new product categories: Virtual Reality (VR) and Dashboard.

Dashboards are defined differently by different authors. Both Few[5] and Wexler et al.[18] emphasize the importance of visualizing information. The difference is that Few[5] highlights the need for information to be quickly and easily understood at a glance, while Wexler et al.[18] includes the use case of monitoring systems. Another definition[18] divides dashboards into functional and visual genres, which highlights the versatility and also emphasizes the interplay between visual and functional elements.

Also, VR has various definitions in the research field and in the understanding of users [6]. For example, Bryson 2013 stated: "Virtual Reality is the use of computer technology to create the effect of an interactive three-dimensional world in which the objects have a sense of spatial presence." [2]. Bryson focuses on computer technology and a three-dimensional world, where Steuer et al. 1995 did not specify this: "A real or simulated environment in which a perceiver experiences telepresence" [16]. Both definitions do not speak of a hedonic system, as it is often seen publicly [20].

The product categories Dashboards as well as VR do not have universally agreed upon definitions. This makes it difficult to derive relevant scales from these definitions. As a result of this paper, we propose relevant scales for these two product categories. The scales we propose were developed through an empirical study that followed the same procedure as the initial study for the product categories [19]. To compare our results with the initial study, we selected a similar product category for each newly proposed category. From our perspective, the most related category to VR is games, as there is a lot of overlap in the development cycle and application scenarios. Dashboards have a lot in common with Learning Platforms (LP), since they are already Plug-Ins in some of them and there must be a similar level of user trust in the displayed data. Therefore, it is to be expected, that a repetition of the original study would produce somewhat similar results for Games and LP, while VR and dashboards would result in comparable scales respectively.

With this background, we derive two research questions (RQ), which we want to answer by addressing the according Hypothesis (H):

RQ 1. *Which scales are relevant for VR and Dashboards for user experience?*

H 1.1. *VR and Games have similar relevant scales for user experience.*

(1) Answers of participants for both product categories correlate.

(2) Answers of participants for both product categories do not correlate with the categories Dashboard and LP

H 1.2. *Dashboards and LP have similar relevant scales for user experience.*

(1) Answers of participants for both product categories correlate.
(2) Answers of participants for both product categories do not correlate with the categories VR and Games

RQ 2. *To what extent is the process of creating new product categories for the UEQ+ of study from Winter et al. [19] reusable?*

H 2.1. *The Original study and our study produce similar results*

(1) LP have similar scales
(2) Games have similar scales

These questions are answered by analyzing the study design of Winter et al. [19] to derive a questionnaire structure for all proposed scales. To check, if our proposed structure produces similar results to the original study (RQ 2), we include two of their product categories, that can be compared to VR and dashboards. To address RQ 2, we use a clustering algorithm in chapter three based on rankings of different sources, including the one from the original source of Winter et al. [19]. Then, we address RQ 1 with quantitative results in the form of top-ranked scales for the new categories VR and dashboards. The section is closed with correlations of product categories, to analyze the hypotheses H 1.1 and H 1.2. Results are discussed and put into perspective with state-of-the-art work in chapter four for both introduced product categories. We conclude this paper by summarizing the key findings in chapter 5.

2 STUDY DESIGN AND REALIZATION

The categories presented in the UEQ+ manual[12] were initially proposed in quantitative studies, and participants of these studies decided which scales are essential. One of these studies is made by Winter et al. [19] and forms the foundation of our study. The procedure to design the questionnaire from Winter et al. was:

(1) create categories based on different software for similar tasks
(2) present these categories in the questionnaire with common known examples
(3) explain each scale with a short sentence
(4) to evaluate the importance build a 7-Likert-scale between -3 (very unimportant) and 3 (very important)

In the following, we give further insights into the construction of the questionnaire in this study and about its implementation.

The questionnaire is constructed with our two new categories *VR* and *Dashboard* as well as two categories from the initial study (*games* and *Learning Platforms*) of Winter et al. [19]. All categories were headlines of a separate page and prominent examples were given for each of them. Categories from the initial study were chosen due to our expectations of high similarity to our newly introduced categories. To minimize participation time, we've decided to select just one similar category for a compact questionnaire [10].

For our study, we adopted all the scales from the initial study[19] and expanded them with scales from the UEQ+[12]. The added scales are: Attractiveness, Haptics, Acoustics, Aesthetics, Intuitive use, and Trustworthiness of Content. Scales that were originally a part of the UEQ+ and designed to measure voice assistance, like Response behavior,

Response quality, and Comprehensibility, were left out. The table 1 shows all 20 scales with a descriptive sentence. Each sentence was used in the questionnaire to give the participant an idea of what the scale is all about.

Name	Abbreviation	Description
Acoustics	Aco	The sound generated by the product is quiet, gentle, and melodious.
Adaptability	Ada	I can customize the product to my personal preferences or work style.
Aesthetics	Aes	I think the product is beautifully designed and attractive.
Attachment	Atm	Even if there are other, similar products for the same tasks, I would not change the product.
Attractiveness	Att	Overall, I think the product is friendly, pleasant, and enjoyable.
Clarity	Cla	I think the product's user interface is clean and clear.
Dependability	Dep	The product always responds to my input in a predictable and consistent way. I am always in complete control of the interactions.
Efficiency	Eff	I am able to achieve my goals with minimal time and effort. The product responds to my inputs quickly.
Haptics	Hap	I find the surface of the product to be stable and pleasant to touch. The surfaces appear to be non-slippery.
Identity	Id	The product enhances my ability to socialize and present myself positively.
Immersion	Imm	I tend to forget about time when I'm engaged with the product. I was completely absorbed in the occupation with the product.
Intuitive use	Int	I am able to operate the product immediately without any training or guidance from others.
Novelty	Nov	The design of the product is unique and interesting. With its original design, it arouses my interest.
Perspicuity	Per	I find the product to be easy to understand and use.
Quality of Content	Qua	The information provided by the product is always up-to-date and of high quality.
Stimulation	Sti	I find the product to be stimulating and exciting. It's fun to deal with.
Trust	Tru	I am confident that my entered data is safe. The data won't be used to hurt me.
Trustworthiness of Content	ToC	The product gives me useful, reliable, and trustworthy information and data.
Usefulness	Use	Using the product brings me benefits. It makes me more productive and saves me time and effort.
Value	Val	I believe the product leaves a professional and high-quality impression.

Table 1. This is a list of all the UEQ+ scales used in the questionnaire, with translated explanations. Originally, the explanations were in German.

The study was conducted with an institutional instance of the online survey tool soSci [1]. The questionnaire consists of seven pages: a welcome text, instructions, four randomly arranged question blocks related to product categories (Dashboard, VR, Games, Learning Platform), and demographic data. Each product category page includes 20 scales, with each scale containing a question that must be answered on the provided Likert scale. The questionnaire concludes by requesting demographic data such as gender, and academic status group. With all questions, the questionnaire contained 83 items and was designed to took 10 to 15 minutes to answer.

[1] https://www.soscisurvey.de/en/index, last visited: 2023-01-04

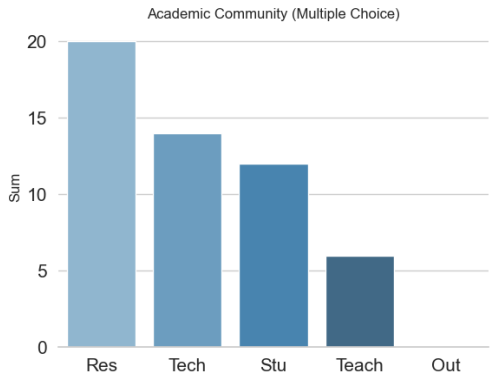
The survey was open to the public from 22-07-05 to 22-11-30. Any German-speaking participant who knows the link to the survey could participate. We have acquired a total of 69 participants via E-Mail lists, personal contacts, students from seminars, and some scientists from Conferences.

3 RESULTS

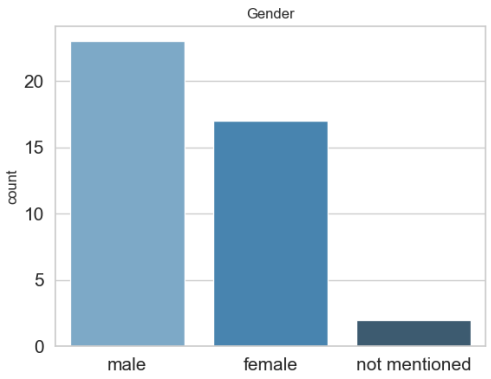
The results of our study are presented in the following section, which is divided into four subsections. An overview of the demographics of our study participants is provided in the first subsection. In the second subsection, our findings are compared to previous studies, with the integration of two established categories into our research included. This enables a comparison of our results to those of previous studies and an examination of the consistency of our findings. The third subsection presents the top scales of our study, highlighting the most important factors that are influential in user experience and satisfaction within each product category. Lastly, in the fourth subsection, the correlation between the product categories is investigated, specifically exploring the proposed relationships between learning platforms and dashboards, as well as games and virtual reality.

3.1 Demographics

From the initial 69 interviews, 24 were filtered out, due to incomplete answers of the questionnaire. Another three were filtered out, due to very fast answers, which indicate a lack of diligence. This leaves our dataset [1] with 42 complete interviews. These 42 interviews were given by 23 male and 17 female participants, while two did not use the option to self report their gender (see Fig. 1b). As one can see in Fig. 1a just people from inside the academic context participated. Furthermore, given our chosen method of communication for acquiring participants, it is reasonable to assume that the majority of them came from the field of computational sciences.



(a) **Academic Community** (Multiple Choice): (**Res**)earch, (**Tech**)nical Staff, (**Stu**)dent, (**Teach**)er, (**Out**)side



(b) **Gender distribution**

Fig. 1. Demographic data

3.2 Clustering Rankings

To address our hypotheses H 2.1(1) and H 2.1(2) we calculated the mean value for each scale in our *study* and ranked them. We also ranked the findings of the first *validation* of UEQ+ [11], and the later extensive *overview*. The *handbook*

of the UEQ+ [12] ranked their found TOP-7 (for games) and TOP-8 (for LP) which we also included in our comparison. Since the ranking in our study included all scales from these three sources, our study counted more scales than any individual source. Thus, the rankings of the individual sources were not given for every scale, which we substituted with NANs (Not a number). To see if our findings relate to the rankings given in the literature, we decided to cluster them. A common approach would be k-means as a clustering algorithm, but since our data contained NANs, we used a modified algorithm, namely kPOD [3]. K in this case equals two, since we assumed two clusters (one for LP and one for games). The algorithm works probabilistically, which is why we iterated it and counted the found clusters. When the iteration count was sufficient to produce a somewhat stable value (500 times), we calculated the mean value of iteration and chosen clusters for each source and category. The result of this algorithm is presented in figure 2.

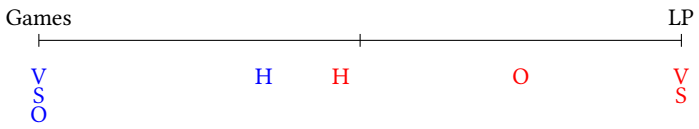


Fig. 2. Comparison of the ranking of scales from **(H)**andbook, this **(S)**tudy, the **(V)**alidation study, an **(O)**verview study using kPOD [3]. Blue letters indicate the ranking for games and red letters are for learning platform(LP).

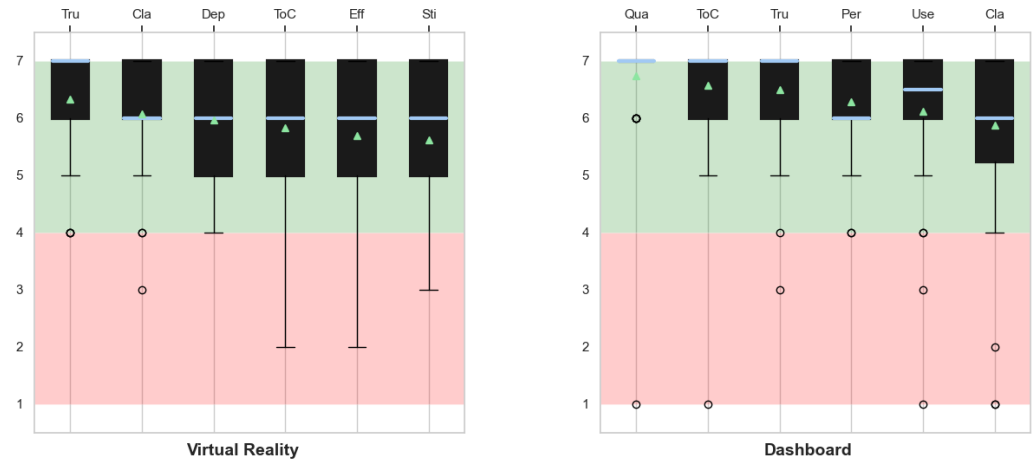
Since the *handbook* just documented the top ranks, there were a lot of NANs, which explains some uncertainty here. The games ranking of the handbook was clustered with other games rankings in 35% of all iterations, while rankings of LP were just clustered with other LP rankings in 47% of all iterations. In most (75%) of all iterations, the ranking of LP in the *overview* was clustered with other LP rankings. The ranking found by our *study* was consistently clustered with rankings of the *validation* ranking in both product categories, and regarding games also with the *overview* ranking.

3.3 Top Scales

Since results of the original categories are comparable with the original study, we can derive relevant scales for our product categories. The ranking of the different scales for VR and dashboards is determined by the mean values. Values rated above 4 on the Likert scale are considered relevant for the product category because they indicate positive importance according to the scale's criteria. All relevant scales for VR are in order: Trust, Clarity, Dependability, Trustworthiness of Content, Efficiency, Stimulation, Perspicuity, Intuitive Use, Acoustics, Immersion, Value, Quality of Content, Attractiveness, Aesthetics, Haptics, Adaptability, Novelty, and Usefulness. Just Attachment and Identity were not considered to be relevant for VR. On the other hand, relevant scales for dashboards were Quality of Content, Trustworthiness of Content, Trust, Perspicuity, Usefulness, Clarity, Efficiency, Intuitive use, Dependability, Value, Adaptability, Attractiveness, and Aesthetics. Scales rated to be not relevant for dashboards were Stimulation, Acoustics, Haptics, Novelty, Attachment, Identity, and Immersion. Due to the high number of relevant scales in both categories (VR: 18; Dashboard: 13), it is necessary to limit the selection of scales to avoid questionnaires of unreasonable length. As suggested from the UEQ+ handbook [12] we decided for the six highest ones (see figure 3).

Having the top-rated scales for the introduced categories (VR and Dashboard) as well as the previous categories (Games and Learning Platforms) lets us check for their similarities. We are doing so by analyzing the correlations between the categories 2.

The correlation between Dashboard and LP indicates an almost identical ranking of importance ratings in all scales (R=0.98), which supports H 1.2(1). Further, the moderate correlations of LP and VR (R=0.53) as well as Dashboards and



(a) Top Scales: **(Tru)**st, **(Cla)**rity, **(Dep)**endability, **(T)**stworthiness **(o)**f **(C)**ontent, **(Eff)**iciency, **(Sti)**mulation;

(b) Top Scales: **(Qua)**lity of Content, **(T)**rustworthiness **(o)**f **(C)**ontent, **(Tru)**st, **(Per)**spicuity, **(Use)**fulness, **(Cla)**rity.

Fig. 3. **Top-rated scales for Virtual Reality and Dashboard**. These box plots show the highest-ranked scales sorted by their mean (light triangle).

dashboard	0.98***		
games	0.05	0.08	
VR	0.53*	0.59**	0.64*
	LP	dashboard	games

Table 2. **Correlation Matrix** of ratings for important scales for **(V)**irtual **(R)**eality **(L)**earning **(P)**latforms, **Dashboards** and **Games**. The color scheme used here represents positive correlations in red and negative correlations in blue, with darker colors indicating stronger correlations. Significance is noted with * (p<0.05), ** (p<0.01) and *** (p<0.001)

.

VR (R=0.59) are somewhat significant and therefore do not support H 1.2(2). The also moderate correlation of VR and Games (R=0.64) shows a quite similar ranking of scales for these two categories, which supports H 1.1(1). Additionally, the missing correlation of Games with either LP (R=0.05) or Dashboards (R=0.08) would support H 1.2(2)., but results for VR do not separate enough from LP (R=0.53) or Dashboards (R=0.59). Even though the strongest correlations appear on the expected comparisons (Dashboard vs. LP and VR vs. Games) with some significance, the according hypotheses H 1.2 and H 1.1 are not fully supported. But the results of checking the hypotheses help us to answer our research question and discuss our findings.

4 DISCUSSION

First, we derive the resulting top scales for the new product categories Dashboard and VR (Image 3). With these results, we answer the research questions stated above, starting with the comparison of our results to the literature, to answer RQ 1. Then, we discuss the derived study design and interpret findings of the clustering and correlation matrix. With these results and found limitations of the study, RQ 2 is answered partially quantitative and partially qualitative.

4.1 Scales for Dashboard

In the presented definitions of dashboards, visualizing information for easy understanding (Few[5]) or monitoring (Wexler et al.[18]) was emphasized. Applying the UEQ+ scales to these definitions suggests that factors such as *trust*, *trustworthiness of content*, *quality of content*, and *usefulness* may be particularly important for the data-related sections. The scales *perspicuity* and *clarity* may be important for ensuring that information is presented in a way that is easy to understand at a glance.

The division into functional and visual genres [18] highlights the versatility of application areas. This suggests that scales such as *aesthetics*, *attractiveness*, *efficiency*, *intuitive use*, and *value* may also be important considerations when designing effective dashboards.

For us, the most important scales to consider when designing dashboards include *perspicuity*, *efficiency*, *quality of content*, *usefulness*, *clarity*, *trust*, and *trustworthiness of content*. The order in which these criteria are listed does not reflect their relative importance, as each may be more or less important depending on the specific context in which the dashboard is being used.

In the presented study, the top five scales for assessing the user experience of dashboards were found to be *quality of content*, *trustworthiness of content*, *trust*, *clarity*, and *usefulness*, which is shown in figure 3b. This aligns with both the definitions and the opinions of the authors, highlighting a consensus on the most important factors to consider when designing effective dashboards. Although the functionalities and applications of dashboards are diverse, there appears to be agreement on the key factors that impact user experience.

The correlation matrix in Table 2 indicates a strong correlation between Dashboard and LP, suggesting that both product categories share similar important scales. This supports the initial idea to have a category from previous studies which is very close to the new proposed one.

4.2 Scales for VR

Definitions of VR usually contain some relation to the sense of presence [6]. Also, there is a fairly long-lasting discussion on which other aspects are crucial for VR [15**?**], which introduces ever more aspects like flow, emotions, place illusion, embodiment, and others. Resulting questionnaires from this discussion are: *Igroup Presence Questionnaire* (IPQ; targeting spatial presence and involvement) [14], *cross-media presence Questionnaire* (ITC; Measuring Sense of Physical Space, Engagement, Ecological Validity, and Negative Effects) [8] or combinations of former questionnaires as the *User Experience in Immersive Virtual Environments* (UXIVE; originally combining 10 aspects in 82 questions from various questionnaires) [17]. These aspects can be vaguely related to the UEQ+ scales, which we interpret with: *Immersion* (from sense of presence and place illusion), *Stimulation* and *Intuitive Use* (from flow), *Visual Aesthetics* and *novelty* (from emotions), and *Haptics* and *Acoustics* (from embodiment).

On the other hand, the public was said to perceive VR as some kind of game. This assumption is further justified by the moderate correlation of VR and games, which would result in a ranking of scales similar to the one of games. The ranking of the UEQ+ handbook includes the following top scales: *Immersion*, *Stimulation*, *Visual Aesthetics*, *Novelty*, *Dependability*, and *Intuitive Use*.

In our study, the most relevant scales were *Trust*, *Clarity*, *Dependability*, *Trustworthiness of Content*, *Efficiency* and *Stimulation*. Therefore, our quantitative findings have very little in common with the otherwise expert-driven approaches of the above-mentioned sources. This can indicate a different understanding of the term VR, which is possible even though we primarily gathered answers from computer scientists. In case, there was a common understanding of VR,

the results depict a point of view compared to experts. While experts seem to focus more on a pleasing experience (e.g. stimulation, visual aesthetics, and novelty), users focus more on pragmatic aspects (e.g. efficiency, dependability, trustworthiness of content, and clarity). This, in return, implies a more hedonic focus of experts, which was attributed above to the perception of the public.

4.3 Adapting the UEQ+ process

As described in section 3.2, our study aimed to produce comparable results for two previously analyzed categories, Games and LP. We decided to run a cluster algorithm, to see if the rankings of all sources can be distinguished from one another. Indeed, most of the iterations clustered our results with the according findings of previous studies. Therefore, it is reasonable to assume that our produced scales are similar to the original study (H 2.1) as well for LP (H 2.1(1)), as for games (H 2.1(2)). Comparing just two product categories and concluding the approach would produce similar scales for every category would be far too broad of an assumption. Therefore, our findings cannot be seen as a proof but as an indicator for our hypotheses and should be set into perspective with other limitations of the studies.

In the following, we will address the issues encountered during the setup, execution, and analysis of the study. While it is impossible to prevent all weak points, it is essential to discuss how to handle these issues and propose possible solutions.

During the study, we encountered several issues, including participant feedback indicating that some had no prior knowledge of the topics and found it challenging to complete the questionnaire. Unfortunately, we did not identify any effective approaches to address this issue, except from motivating every participant that their opinion is important and that expertise only helps to a limited extent. Another point was that some phrasing in the questions was viewed as too complicated, which could have caused confusion among participants. Due to our commitment to remain as close to the initial study as possible, we had limited flexibility in modifying the question phrasing to make them more accessible. We also found that the use of four categories in the questionnaire may have been excessive, although it was difficult to strike a balance between the information gained from the study and the time commitment required. However, the average participation time of round about 9 minutes was not exceeded, which is an acceptable length for a questionnaire[10].

The study's replicability was weakened by the fact that the initial study was conducted in Excel, and the document was inaccessible. To improve this, the study doesn't have to be connected to proprietary software. In our supplementary [1], the data is open in a CSV file with data points, possible questionnaire values and variables. In any further studies, it could be helpful to have integrable data. Having just some participants, as suggested in the initial study, can not level out different expectations of use cases for the product categories, and can therefore produce unexpected results. But this open data could help address the issue of the small sample size. The relatively homogenous participant field, consisting mainly of academics from the field of computer science, could have been beneficial for a common understanding of terms, but makes the results less transferable for society. Further studies could aim to expand the participant pool and increase diversity.

This discrepancy in importance of scales for VR could be a result of the vast development of the technology in recent years. A few years back, it was reasonable to set Games and VR on one level. Nowadays, the usage of such technology reached into social interaction in conferences, offers various simulations for educational purposes, is used in psychotherapy and is a part of advertisement when visiting your new hotel virtually, while still maintaining the possibility to play some games.

5 CONCLUSION

In this article, new product categories, namely Dashboard and Virtual Reality, were formed for the UEQ+ based on the schema from previous studies. The important scales were formed by a study, in which participants were presented with all scales from the UEQ+ and asked to determine which scales were important from their perspective. In order to validate the results and compare them with previous studies, the importance of the scales for the similar categories of Learning Platform and Games was also determined. The analysis indicated that participants found similar scales important to those already presented, which are:

Dashboard *Quality of Content*, *Trustworthiness of Content*, *Trust*, *Clarity*, and *Usefulness*

Virtual Reality *Trust*, *Clarity*, *Dependability*, *Trustworthiness of Content*, *Efficiency* and *Stimulation*.

The raw data can be viewed in the Open Science Framework data repository[1], and all analytical scripts are available in the src-folder, including explanations provided in the Jupyter Notebook. This study demonstrates that the process for creating new product categories is reproducible and allows for a low entry barrier for further formations of product categories due to the open accessibility of all data.

REFERENCES

[1] anomynous1 and anomynous2. 2023. Blinded for review. osf.io/?????

[2] Steve Bryson. 2013. Virtual Reality: A Definition History - A Personal Essay. (Dec 2013). https://doi.org/10.48550/arXiv.1312.4322 arXiv:1312.4322 [cs].

[3] Jocelyn T. Chi, Eric C. Chi, and Richard G. Baraniuk. 2016. k-POD: A Method for k-Means Clustering of Missing Data. *The American Statistician* 70, 1 (Jan. 2016), 91–99. https://doi.org/10.1080/00031305.2015.1086685 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00031305.2015.1086685.

[4] Fred D. Davis. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 3 (1989), 319. https://doi.org/10.2307/249008

[5] Stephen Few. 2006. *Information Dashboard Design* (1. ed.). O'Reilly.

[6] Suzan (Suzie) Kardong-Edgren, Sharon L. Farra, Guillaume Alinier, and H. Michael Young. 2019. A Call to Unify Definitions of Virtual Reality. *Clinical Simulation in Nursing* 31 (Jun 2019), 28–34. https://doi.org/10.1016/j.ecns.2019.02.006

[7] Bettina Laugwitz, Martin Schrepp, and Theo Held. 2006. *Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten.* Oldenbourg Wissenschaftsverlag, 125–134. https://doi.org/10.1524/9783486841749.125

[8] Jane Lessiter, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. 2001. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments* 10, 3 (Jun 2001), 282–297. https://doi.org/10.1162/105474601300343612

[9] John Long and John Dowell. 1990. Conceptions of the discipline of HCI: craft, applied science, and engineering. In *Proceedings of the fifth conference of the British Computer Society, Human-Computer Interaction Specialist Group on People and computers V*. Cambridge University Press, USA, 9–32.

[10] A. Ant Ozok. 2007. *Survey Design And Implementation in HCI* (2nd ed.). CRC Press, Boca Raton, 1384. https://doi.org/10.1201/9781410615862

[11] Martin Schrepp and Jörg Thomaschewski. 2019. Design and Validation of a Framework for the Creation of User Experience Questionnaires. *International Journal of Interactive Multimedia and Artificial Intelligence* 5, Regular Issue (2019). https://www.ijimai.org/journal/bibcite/reference/2730

[12] Martin Schrepp and Jörg Thomaschewski. 2020. Handbook for the modular extension of the User Experience Questionnaire. (Jul 2020). https://ueqplus.ueq-research.org/Material/UEQ+_Handbook_V2.pdf

[13] Martin Schrepp, Jörg Thomaschewski, and Andreas Hinderks. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, Regular Issue (2017). https://www.ijimai.org/journal/bibcite/reference/2604

[14] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (Jun 2001), 266–281. https://doi.org/10.1162/105474601300343603

[15] Kay M. Stanney, Ronald R. Mourant, and Robert S. Kennedy. 1998. Human Factors Issues in Virtual Environments: A Review of the Literature. *Presence* 7, 4 (Aug 1998), 327–351. https://doi.org/10.1162/105474698565767 Conference Name: Presence.

[16] Jonathan Steuer, Frank Biocca, Mark R Levy, et al. 1995. Defining virtual reality: Dimensions determining telepresence. *Communication in the age of virtual reality* 33 (1995), 37–39.

[17] Katy Tcha-Tokey, Emilie Loup-Escande, Olivier Christmann, and Simon Richir. 2016. A questionnaire to measure the user experience in immersive virtual environments. In *Proceedings of the 2016 Virtual Reality International Conference (VRIC '16)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/2927929.2927955

[18] Steve Wexler, Jeffrey Shaffer, and Andy Cotgreave. 2017. *The big book of dashboards: visualizing your data using real-world business scenarios.* Wiley, Hoboken, New Jersey.

[19] Dominique Winter, Andreas Hinderks, Martin Schrepp, and Jörg Thomaschewski. 2017. Welche UX Faktoren sind für mein Produkt wichtig? (2017). https://doi.org/10.18420/muc2017-up-0002

[20] Raphael Zender, Josef Buchner, Caterina Schäfer, David Wiesche, Kathrin Kelly, and Ludger Tüshaus. 2022. Virtual Reality für Schüler:innen: Ein «Beipackzettel» für die Durchführung immersiver Lernszenarien im schulischen Kontext. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung* 47 (Apr 2022), 26–52. https://doi.org/10.21240/mpaed/47/2022.04.02.X