Assessing User Apprehensions About Mixed Reality Artifacts and Applications: The Mixed Reality Concerns (MRC) Questionnaire

Christopher Katins
christopher.katins@hu-berlin.de
HU Berlin
Berlin, Germany

Paweł W. Woźniak
pawel.wozniak@chalmers.se
Chalmers University of Technology
Gothenburg, Sweden

Aodi Chen
aodi.chen@student.hu-berlin.de
HU Berlin
Berlin, Germany

Ihsan Tumay
ihsan.tumay@hu-berlin.de
HU Berlin
Berlin, Germany

Luu Viet Trinh Le
leluvitr@hu-berlin.de
HU Berlin
Berlin, Germany

John Uschold
john.uschold@student.hu-berlin.de
HU Berlin
Berlin, Germany

Thomas Kosch
thomas.kosch@hu-berlin.de
HU Berlin
Berlin, Germany

ABSTRACT

Current research in Mixed Reality (MR) presents a wide range of novel use cases for blending virtual elements with the real world. This yet-to-be-ubiquitous technology challenges how users currently work and interact with digital content. While offering many potential advantages, MR technologies introduce new security, safety, and privacy challenges. Thus, it is relevant to understand users' apprehensions towards MR technologies, ranging from security concerns to social acceptance. To address this challenge, we present the Mixed Reality Concerns (MRC) Questionnaire, designed to assess users' concerns towards MR artifacts and applications systematically. The development followed a structured process considering previous work, expert interviews, iterative refinements, and confirmatory tests to analytically validate the questionnaire. The MRC Questionnaire offers a new method of assessing users' critical opinions to compare and assess novel MR artifacts and applications regarding security, privacy, social implications, and trust.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**.

KEYWORDS

Mixed Reality, User Apprehensions, Concerns, Privacy, Safety, Security, Social Acceptance, Trust

1 INTRODUCTION

Mixed Reality (MR) [70] research is a growing field covering a broad spectrum of technologies and applications that blur the boundaries between digital and real worlds. Considering the evolution of MR over the past years, we observed that many innovations have primarily brought incremental improvements to MR technologies. As a consequence, MR devices become more commonly available through smartphones [50] and even more interwoven by using head-mounted displays [45, 54]. Fueled by the commercial success of the Microsoft HoloLens 2[1] in industry settings and further expectations towards the Apple Vision Pro[2] in consumer use, MR might become omnipresent soon.

The transition from fiction to reality has steadily progressed in recent decades with the continued research in this field and the emergence of commercially available MR products. Previous research has extensively investigated use cases (e.g., in the context of work [56] or education [35]) as these technologies become increasingly accessible. At the same time, evaluating their usability and potential benefits is essential, as well as understanding the concerns and apprehensions that MR devices raise with their integration into our lives. Existing issues related to hardware performance, software optimization, and interaction design tend to improve over time as computational power increases and hardware shrinks. As a result, technical challenges that currently hinder seamless MR experiences will likely diminish over time. Yet, it is essential to recognize that the evolution of MR is not solely a matter of technological advancement. The challenge lies in addressing individuals' potential apprehensiveness about the technology.

---

[1]https://www.microsoft.com/en-us/hololens, last accessed on 2023-12-12.
[2]https://www.apple.com/apple-vision-pro, last accessed on 2023-12-12.

In this context, a new challenge emerges for HCI: Understanding how individuals apprehend novel MR systems regarding their perceived concerns about this technology. Numerous methodologies and tools have been developed for evaluating user experience (e.g., UEQ [66]), usability (e.g., SUS [26]), or acceptance (e.g., TAM [18]). At the same time, researchers have rarely investigated users' apprehensions and concerns regarding novel MR technologies besides usability measures. Thus, measuring user apprehensions and concerns remains a research gap.

This paper presents the Mixed Reality Concern (MRC) Questionnaire to address these challenges. The MRC enables an evaluation that extends beyond usability and other aspects of the new system, encompassing potential concerns and apprehensions. Our systematic approach to developing this scale was based on the guidelines by Boateng et al. [3]. Initially, a conceptual model of potential concerns was formulated, drawing from relevant research in the field. This model comprises four primary categories, with 30 subcategories that cover a broad spectrum of potential user concerns. The model is shown in Table 1. Subsequently, an initial set of 120 items derived from this conceptual model was generated. These items were then refined through expert feedback and underwent an exploratory factor analysis, resulting in the final scale composed of 9 items. A comprehensive evaluation of this scale followed to ensure the validity of its results. Finally, we anticipate the MRC to complement current usability metrics by acting as a tool for researchers and practitioners to measure concerns towards their MR applications and artifacts. Given that MR is a technology distinct from traditional user interfaces and devices that increasingly proliferates into home and work environments [70], users may assume implicit or explicit concerns that significantly influence their interaction with these artifacts. The questionnaire is designed to concentrate specifically on MR-related user concerns, facilitating practitioners in quickly and comprehensively understanding potential apprehensions that could affect the overall user experience.

2 RELATED WORK

With the rapid advancements in MR technology, understanding users' apprehensions about MR technology is crucial for its successful integration into everyday life. MR has shown tremendous potential in various domains, but its widespread adoption is impeded by several challenges that need to be addressed for it to become a mainstream technology [35]. By giving an overview of current research about novel challenges in MR, we aim to provide a comprehensive backdrop against which user concerns can be effectively evaluated in the later sections.

2.1 Social Acceptance and Social Implications: Challenges to the Ubiquity of MR

One of the critical barriers to the widespread acceptance of MR is the lack of social acceptance. A 2021 study by Thomas et al. [71] sheds light on the barriers to social acceptance surrounding MR devices. Despite the functional benefits of MR technology, the study reveals that certain facts genuinely worry everyday users. One of the primary barriers is the perceived social awkwardness associated with wearing MR devices in public, which can lead to feelings of self-consciousness and reluctance to embrace this technology.

Moreover, the study mentions that the appearance and design of MR devices are critical factors influencing social acceptance, as aesthetically unappealing or intrusive devices may deter individuals from incorporating them into their daily lives. To foster broader social acceptance of MR, the study emphasizes the importance of improving the functionality and user experience and addressing these social and psychological barriers to ensure MR devices become seamlessly integrated into society's fabric.

Slater et al. [69] determined a number of ethical considerations that ought to be considered in the future development of MR technologies. Next to common privacy concerns due to the vast amount of data collected by MR devices (further discussed in Section 2.2), the publication illustrates how highly realistic VR and AR environments can impact users emotionally, psychologically, and socially. These impacts include but are not limited to the ubiquity of MR, akin to mobile technology, as it can impede meaningful real-world interactions, potentially resulting in social isolation. This shift towards MR may also cultivate a preference for virtual interactions over real-life ones, leading to societal withdrawal. Moreover, the potential "superrealism" of MR experiences may lead some individuals to neglect their physical well-being, paralleling extreme cases of excessive video game usage where the boundary between the virtual and physical worlds blurs. Immersive MR environments can also encourage imitative behaviors that individuals would typically avoid in reality, either through gradual exposure or emulation of actions taken by virtual characters. The persuasive power of MR, particularly in highly realistic iterations, raises ethical concerns when employed to modify emotions and behaviors for potentially harmful ends. Furthermore, this capacity to manipulate sensory experiences raises questions about the reliability of sensory evidence in both legal and societal contexts.

2.2 Security, Safety, and Privacy: Common Threats in a New Environment

According to Gugenheimer et al. [27], while a significant portion of research focuses on technological advancements in MR, it is equally crucial to emphasize research into the potential hazards and challenges that accompany these innovations. They determined the well-established topics of security, safety, and privacy in general computer science to be relevant for the MR research. These aspects become more important since they proliferate into other research areas for wider adoption, including MR support at production lines [6, 58], education [23, 47], or transportation [41, 42, 49] while changing the perception and interaction capacities of users [22, 67, 70]. With such growth in MR, privacy concerns encompass two main viewpoints: that of the user and that of bystanders. User-related privacy issues revolve around the risks associated with biometric identification or surveillance of behavior and attention. In contrast, bystander privacy concerns how MR sensors, such as cameras, may impact individuals who did not consent to be observed by the technology [14].

In the context of trust, Jian et al. [38] discuss the increasing prevalence of automation in complex systems and everyday life. The authors review existing research on measuring trust in various contexts, such as social psychology and human-machine systems, highlighting the multidimensional nature of trust and the need for

a more empirical understanding. Furthermore, the authors identify and scrutinize previous studies, including the lack of differentiation between trust and distrust, and emphasize the importance of assessing trust in the context of human-machine systems, leading to the necessity for the development of an empirically based tool for assessing trust in increasingly automated environments.

Further, Harborth and Pape [30] also report that technical assessments of risks related to MR reveal that the technology introduces new privacy concerns that require immediate attention. Individuals using MR genuinely worry about their privacy, and these apprehensions significantly deter technology adoption. The study highlights the importance of addressing these privacy risks promptly and effectively to foster trust and confidence among users.

A unique aspect emerging in MR research is "immersive attacks," [1, 8, 76] which target users' physiological and psychological safety through perceptual manipulation rather than exploiting hardware or software vulnerabilities. These attacks leverage perceptional illusions and necessitate the development of protective layers to detect and prevent such manipulations, highlighting the distinctive challenges posed by MR technology.

Lastly, safety and health concerns are yet another barrier that must be addressed to facilitate the broader adoption of MR. Yuntao Guo et al. [28] reported on the safety and health concerns associated with location-based MR gaming applications. As these games blur the lines between virtual and physical environments, potential risks and hazards emerge that can impact players' well-being. The study mentions that one primary concern is the distraction factor, where players may become engrossed in the game and fail to pay adequate attention to their surroundings, leading to accidents or injuries. Additionally, prolonged usage of MR gaming apps can result in physical strain, eye discomfort, and even musculoskeletal issues, especially when players engage in prolonged or repetitive gameplay [43]. The study emphasizes the importance of understanding these safety and health implications, particularly for game developers and policymakers, to implement safety measures, provide user guidelines, and raise players' awareness of the responsible use of location-based MR gaming apps.

2.3 Related Questionnaires

Next to the objective key challenges that pertain to MR, acquiring the feedback of users is invariably a crucial part of the development of new technologies, be it in the field of MR or elsewhere. To this end, numerous questionnaires and scales have been developed to assess various aspects of user experiences within this domain. However, it is essential to note that these existing questionnaires often focus on specific dimensions of user perceptions and do not comprehensively address the diverse spectrum of concerns that may arise. This section briefly overviews these related questionnaires, highlighting their strengths and limitations.

One of the most widely known measures of user acceptance of technology is the **Technology Acceptance Model (TAM)**, developed by Fred Davis in the 1980s [17, 18]. It aims to measure acceptance by determining both the *ease of use* and the *perceived usefulness* of a technological system. The TAM has been further developed [72, 73], and other publications aimed at extending the model by adding further factors, such as *perceived enjoyment* [51, 63].

Notable is also the **Attitudes toward Virtual Reality Technology Scale (AVRTS)** [5], using the TAM as an initial model to then further develop a scale to assess attitudes towards VR technologies. All in all, the TAM, its variants, the AVRTS, and other commonly used scales in HCI research like the **System Usability Scale (SUS)** [26], the **AttrakDiff** [31], or the **User Experience Questionnaire (UEQ)** [66] is based on assessing the acceptance, the general usability, the hedonic and pragmatic qualities, and the general user experience respectively. While these scales excel at evaluating usability and gauging user affinity for a particular artifact, their design does not prioritize the measurement of concerns or unfavorable opinions regarding those devices.

The **Perceived Creepiness Technology Scale (PCTS)** [78] stands out in this respect as it specifically seeks to evaluate an adverse emotion. The primary purpose of the **PCTS** is to allow designers and researchers to quickly assess new technologies that might elicit initial sensations of creepiness in users in that regard.

Next to the **AVRTS**, scales like the **Augmented Reality Immersion (ARI)** questionnaire [25] and various presence questionnaires [62, 77] seek to ensure that the measurements are relevant and accurate when applied to MR use cases, necessitating the development of novel questionnaires tailored to these technologies. The **Virtual reality sickness questionnaire (VRSQ)** [44] and the **Augmented Reality Sickness Questionnaire (ARSQ)** [36] aim to measure the immediate negative impact of MR on the users' well-being, but to the authors' best knowledge, no scales exist that aim to determine the long-term effect of MR on its users.

Remotely related is the Concerns-Based Adoption Model (CBAM) with its **Stages of Concern Questionnaire (SoCQ)** [24], an educational framework developed in the late 1970s. It is designed to understand and facilitate the process of educational innovation and change, particularly in the context of school settings. Although the questionnaire may not be suitable for assessing concerns related to MR technology and its users, the stages it outlines provide valuable insights into how individuals perceive innovations and their potential reactions to them.

3 CONCEPTUAL FRAMEWORK: CATEGORIZING CONCERNS ABOUT MR

Based on the findings of Section 2, a preliminary conceptual framework was developed to categorize potential user concerns about MR systems. As this classification is derived from related literature, it can logically only serve as a framework for classifying the ongoing research within this domain. Acknowledging that such categorizations may not always align with users' subjective concerns or considerations is essential. Hence, this only represents an initial basis from which the subsequent construction of the scale could proceed as further explained in Section 4.

The decision to develop a preliminary conceptual framework for generating the questionnaire items rather than to base it on psychological models, such as the Innovation Resistance Theory (IRT) [60], was driven by the recognition that possible concerns regarding MR might extend beyond the generic barriers that are often defined for novel technologies or innovations as a whole. Herein, contemporary issues such as privacy, which are crucial in the field of MR, are often only implicitly addressed in existing
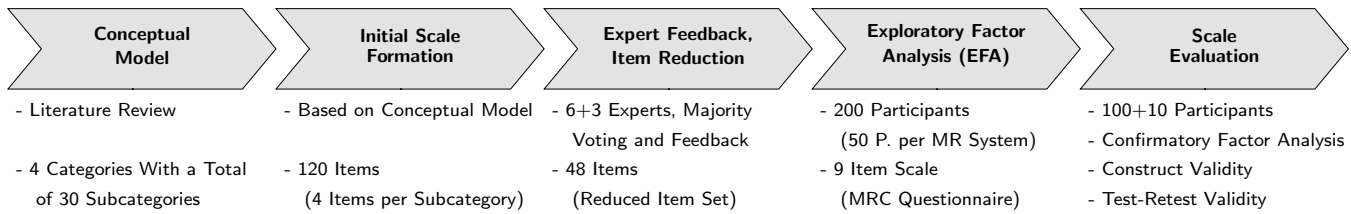
Conceptual Model	Initial Scale Formation	Expert Feedback, Item Reduction	Exploratory Factor Analysis (EFA)	Scale Evaluation
- Literature Review	- Based on Conceptual Model	- 6+3 Experts, Majority Voting and Feedback	- 200 Participants (50 P. per MR System)	- 100+10 Participants
- 4 Categories With a Total of 30 Subcategories	- 120 Items (4 Items per Subcategory)	- 48 Items (Reduced Item Set)	- 9 Item Scale (MRC Questionnaire)	- Confirmatory Factor Analysis - Construct Validity - Test-Retest Validity

Figure 1: The process of developing the scale, as this paper outlines.

models, if at all. Hence, deriving potential concerns from currently recognized challenges in MR was deemed more fitting, ensuring that the questionnaire reflects the nuanced research field of MR and addresses issues that may not be adequately captured by existing psychological models.

3.1 Security and Privacy: Contrasting, yet not Mutually Exclusive

The categorization of security threats in MR is based on the publication "Security and Privacy Approaches in Mixed Reality: A Literature Survey" [19]. It compiles various strategies suggested to maintain the security and privacy of users and data within the realm of MR in previous work. Furthermore, the researchers combined the already existing security and privacy properties from previous work [20, 34, 40] for a final scale consisting of six security-related properties and six privacy-related properties on each end, with one property being related to both. They observed that specific security attributes may be simultaneously perceived as potential privacy risks. They note that this underscores the variations in the emphasis placed on these attributes or prerequisites by different stakeholders.

This categorization provides a comprehensive overview of the security and privacy risks in MR that are presently recognized in research and actively addressed, conceivably also covering the concerns that users of MR systems might have in this regard. As a result, the aforementioned properties form two of the four principal categories within our framework.

3.2 Social Implications: Psychological Safety, Health, and Social Impact

Safety, specifically psychological safety, is another novel challenge in MR [27]. In this context, the publication "The Ethics of Realism in Virtual and Augmented Reality" [69] identified eleven potential psychological and social implications that should be considered in the future development of MR. Given the extensive range of potential social impacts, achieving comprehensive coverage is unattainable. Yet, to consider a broad range of potential psychological and social concerns, we chose to integrate each implication as a subcategory under the respective factor.

3.3 Public Acceptance: Perception and Trust

Numerous factors can potentially shape the public's willingness to embrace novel technologies. The publication "Socio-psychological determinants of public acceptance of technologies: A review" [29]

sought to explore the psychological factors that underlie the societal acceptance of emerging technologies and assembled a list of the most frequently employed determinants found in related research. We opted for choosing a subset of these determinants that seemed fitting for the application regarding MR technology, especially considering the findings of Section 2.1. Herein, the primary emphasis centers on the perception of the technology rather than its actual properties and the level of trust in these systems.

4 SCALE FORMATION

After establishing a related-work-based initial conceptual framework for categorizing potential user concerns about MR, the subsequent phase involved developing a questionnaire that covers the genuine apprehensions of users. We followed a systematic procedure to accomplish this, as illustrated in Figure 1. This procedure is based on the scale development best practices proposed by Boateng et al. [3]. This approach closely aligns with the methodology employed for developing the PCTS [78], which also aims to capture critical sentiments regarding novel technologies.

4.1 Item Generation

The initial items were generated by two researchers, creating four items for each subcategory of the conceptual framework, resulting in a total of 120 items. As the related work [19, 29, 69] gave definitions for each property/implication, we generated similar, albeit slightly different phrasing to allow for a more nuanced set of items in the end. Afterward, the authors discussed the items and revised items that sounded too similar.

4.2 Expert Feedback

Two rounds of expert feedback were carried out to reduce the substantial pool of initial items. In the first round, six experts were asked to give feedback on the initial set of items and indicate whether they considered each item essential for such a scale. The experts were researchers in the fields of privacy, security, VR/AR, and general HCI.

The reduced set of items was chosen through majority voting, meaning that only if at least three experts indicated an item to be essential, it was retained, and all other items were discarded. The remaining items were then discussed and improved upon by the researchers based on the initial feedback of the experts. This resulted in a final set of 48 items.

Subsequently, another round of expert feedback was gathered for a final iteration, specifically regarding the phrasing to ensure that all items are easily comprehendable and sufficiently distinct. The

Table 1: The preliminary conceptual framework with its four categories and their respective subcategories aiming to classify potential user concerns regarding MR systems. This model will be used to develop the scale in the following.

User Concerns About MR Systems
Security [19]
Integrity
Non-Repudiation
Availability
Authorization
Authentification
Identification
Confidentiality
Privacy [19]
Anonymity & Pseudonymity
Unlinkability
Unobservability & Undetectability
Plausible Deniability
Content Awareness
Policy & Consent Compliance
Social Implications [69]
Social Isolation
Preference for Virtual Social Interactions
Body Neglect
Imitative Behavior
Persuasion
Unexpected Horror
Pornography and Exposure to Violence
Extreme Violence and Assault
Lack of Common Environments
Lack of Ground Truth
Persuasive Advertising
Public Acceptance [29]
Perceived Health Implications
Social Outcast
Interactions
Trust
Family & Friends
Perceived Risk

three experts involved in the second round differed from those who participated in the initial round. They are researchers in the fields of VR/AR, human augmentation, and general HCI, respectively. Two of the experts had previous experience in developing questionnaires, while one expert, although knowledgeable about the process, had not previously engaged in questionnaire development. To maintain balance and minimize potentially leading questions,

half of the items in the final set were reversed. This was done to ensure that overwhelmingly negative phrasing would not skew responses, reducing bias where possible.

4.3 First Survey

After developing the reduced set of initial scale items, based on related work and expert feedback, a participant study was executed to refine the item set further through exploratory factor analysis. In accordance with the sample size recommendation by Comrey [13], $n = 200$ participants were recruited.

4.3.1 Participants. Prolific[3] was used to recruit participants, ensuring a more representative sample of subjects than through institute mailing lists or similar approaches. The participation was entirely voluntary, and the option to withdraw from the survey was available throughout. Participants were compensated with £1.50 upon completing the survey, corresponding to an average hourly reward of £15.15. The survey was conducted entirely online and took approximately 10 minutes to complete. The average age of participants was roughly 40 years ($\bar{x} = 39.65$, $s = 12.94$), 50% identifying as male, 50% identifying as female, and all either currently residing in the United Kingdom or the United States.

4.3.2 Survey Structure. To verify the robustness of our model in representing user concerns across various implementations of MR, four versions of the survey were created, with two versions introducing an AR prototype and two versions showing a VR prototype instead. Each version was shown to 25% of the participant pool, ensuring equal distribution. Furthermore, one prototype per technology was described to feature functionality that is usually linked to be rather concerning, while the other prototype was selected to showcase features that are typically associated with lower levels of concern. This was done to ensure the scale could consistently gauge concerns across a spectrum of intensities for various types of MR technologies. They were each described with neutrally phrased text of roughly 200 to 300 words and a mockup image of the interface/system. The participants were asked to state how much they agreed with each item of the reduced item set on a 5-item Likert scale (Strongly disagree, Disagree, Neutral, Agree, Strongly agree).

All four prototypes were based on related work and already existing technologies. The non-concerning AR system was an intelligent navigation system, showing navigational clues via holograms and rerouting the user based on their preferences and current traffic information. This is based on already existing systems, implemented and tested in both research and industry environments [2, 57]. The concerning AR system was based on "FlirtAR"[4] and "ARR, matey!"[5], describing a dating app that would show information about the conversation partner and conversational suggestions via AR. The non-concerning VR system featured a virtual vacation application, similar to a multitude of readily available VR apps[6] and related research [53, 59]. Lastly, the concerning VR prototype featured a gaming scenario, which would adapt the difficulty based on the

---

[3]https://www.prolific.co, last accessed on 2023-12-12.

[4]https://flirtar.co, last accessed on 2023-12-12.

[5]https://wp.nyu.edu/tlt/conversational-roleplay-using-augmented-reality/, last accessed on 2023-12-12.

[6]https://www.meta.com/de-de/blog/quest/virtual-vacation-11-vr-apps-and-films-that-let-you-travel-the-world-from-home/, last accessed on 2023-12-12.

player's emotions and physiological signals, porting the preexisting work of Chanel et al. [10] into a VR environment.

4.4 Exploratory Factor Analysis

Analogous to the development of other scales in the field of HCI [55, 75, 78], the extraction of latent factors was conducted as proposed by McCoach et al. [52]. The results of the reversed items were inverted, and the Kaiser-Meyer-Olkin (KMO) criterion [68] was evaluated. With KMO values above 0.8 indicating satisfactory sampling adequacy and the result being KMO = 0.93 for the present dataset, we continued with the factor analysis. For this, the parallel factors technique [33] was used in conjunction with a Scree plot [9] to find the optimal number of principal axis factors. A varimax rotation was applied, as this orthogonal rotation method produces independent factors, aiming to allow the later reduction of items that load on multiple factors at once [79]. Herein, the scree plot analysis indicated three factors to be the optimal solution for the items at hand.

To further reduce the set of items to achieve a concise scale that is practical for application in MR research, items with factor loadings below 0.40 were removed, as they are generally considered inadequate for such models [61]. Items with significant cross-loadings were consequently removed as well. The final scale consists of 3 items per factor, leading to a number of 9 items in total. Cronbach's alphas, indicating the internal consistency of the (sub)scales, all show adequate consistency for the three factors, and the overall Cronbach's alpha of $\alpha = 0.85$ for the scale as a whole confirms that suitable items were retained [15]. The Cronbach's alphas of the subscale and the factor loadings of the items are all shown in Table 2. The model displays a good fit with KMO = 0.81, a Tucker Lewis Index [7] of TLI = 0.98 and a Root Mean Square Error of Approximation of RMSEA = 0.049.

4.4.1 Factor Naming. As the first three items (**SP1**, **SP2**, **SP3**) are a combination of the two subcategories *Security* and *Privacy*, we opted to name the first factor **Security & Privacy**. This is in accordance with the work of De Guzman et al. [19], as introduced in Section 3.1, where the two factors were also combined into one contiguous list of properties. The items **SI1**, **SI2**, and **SI3** all stem from the **Social Implications** subcategory of the conceptual framework, making the naming of the second factor trivial. Interestingly, the last three items (**T1**, **T2**, **T3**) all were reversed items. While their content in part correlates with the *Security* and *Privacy* categories, they also closely align with the last category, that being *Public Acceptance*, or to be more precise, *Trust*. As other properties of the *Public Acceptance* are not present in the final item set anymore and with the first factor already covering the potential concerns regarding both *Security* and *Privacy*, we decided to name this factor **Trust**. With this factor only consisting of reversed items, we hope also to reduce the latent negative bias that might stem from the critically phrased items of the preceding factors.

5 SCALE EVALUATION

With the three final factors of the scale being determined, the MRC Questionnaire could now be evaluated appropriately. This process followed the *Phase 3* of scale development by Boateng et al. [3] and included two further surveys.

5.1 Second Survey

The first of the two surveys for evaluation was carried out to gather data for a confirmatory factor analysis, convergent/divergent validity, and differentiation by known groups.

5.1.1 Participants. As with the first survey, see Section 4.3, we again chose to use Prolific as the recruitment platform for this survey. Similarly, the participation was entirely voluntary, and the option to withdraw from the survey was available throughout. Participants were compensated with £1 upon completion of the survey, which corresponds to an average hourly reward of £13.62. In total, $n = 100$ participants were recruited for this survey. It was conducted entirely online and took approximately 5 minutes to complete. The average age of participants was again roughly 40 years ($\bar{x} = 39.83$, $s = 12.05$), 50% identifying as male, 50% identifying as female, and all either currently residing in the United Kingdom or the United States.

5.1.2 Survey Structure. Participants were shown one of two prototypes for assessment, one again being expected to yield comparatively few concerns and one raising potentially more concerns. Each was depicted using neutral-worded descriptions, spanning approximately 200 to 300 words, along with a mockup image of the interface/system. Each participant was randomly assigned one of the two prototypes, and both were shown with equal frequency.

Both systems offered the same fundamental feature set, namely an AR application offering contextual information for tourists in cities unfamiliar to them. This included the navigation to relevant points of interest through holograms that blend into the environment for unobtrusive clues, offering an adaptive AR experience. The second prototype introduced an additional feature, specifically blocking the view of parts of reality based on user preference. The hypothetical adaptive AR base system is based on related work [16, 39], and the added view filter has been discussed in recent publications [21] and expert interviews[7] as well.

After an introduction to the prototype at hand, participants were instructed to state their agreement with each of the items of the final MRC Questionnaire as shown in Table 2. Additionally, they were asked to complete both the PCTS [78] and the UEQ [66] for the shown prototype to facilitate convergent/divergent validity tests.

5.2 Confirmatory Factor Analysis

To evaluate the structural validity of the scale, we performed a Confirmatory Factor Analysis (CFA). Herein, the dimensionality of the model can be verified through systematic fit assessments, confirming the structure of the model if certain thresholds are met [3]. With a Tucker Lewis Index of TLI = 0.98, a Comparative Fit Index of CFI = 0.99, and a Root Mean Square Error of Approximation of RMSEA = 0.059 the results are indicative of an internally consistent model with a fair to close fit. As seen in Figure 2, the subscales exhibit a moderate to high correlation, implying that the theoretical scale is reasonable. The Cronbach's alphas for the three subscales are $\alpha = 0.92$ for *Security & Privacy*, $\alpha = 0.85$ for *Social Implications*, and $\alpha = 0.79$ for *Trust*, respectively.

---

[7]https://www.businessinsider.com/facebook-meta-metaverse-splinter-reality-more-2021-11, last accessed on 2023-12-12.

Table 2: The final MRC Questionnaire, comprised of three factors with three items each. Also reported are Cronbach's alphas and factor loadings based on the first survey results.

Subscale/Item	ID	Factor Loading		
		SP	SI	T
Security & Privacy, $\alpha = 0.88$				
I am concerned about the possibility of non-authenticated individuals gaining access to this MR system.	**SP1**	0.80		
I am concerned about the potential exposure of sensitive data through this MR system to unauthorized parties.	**SP2**	0.80		
I worry that using this MR system might lead to my personal information being misused.	**SP3**	0.80		
Social Implications, $\alpha = 0.81$				
I fear that with this MR system, it becomes increasingly hard to maintain a clear distinction between virtual behavior and real-life behavior.	**SI1**		0.78	
I am concerned about the potential of this MR system to influence my behaviors in ways that could be detrimental to my well-being.	**SI2**		0.75	
Using this MR system might make me appear disconnected from others in my physical environment.	**SI3**		0.74	
Trust, $\alpha = 0.88$				
I believe that only legitimate individuals can access this MR system. (R)	**T1**			0.77
I am sure that this MR system is maintaining a secure environment. (R)	**T2**			0.78
I am confident that my anonymity is protected by this MR system. (R)	**T3**			0.74

5.3 Construct Validity

As the two prototypes for the second survey were consciously chosen to differ in the number of concerns raised through having the same set of basic features, but the second one added a real-life filter that is already critically discussed in current literature, a differentiation by known groups is possible. Afterward, the MRC Questionnaire is compared to existing scales to evaluate if and how different concepts correlate with the proposed model.

5.3.1 Differentiation by known groups. The two prototypes for the second survey were intentionally selected to raise varying levels of concerns by sharing the same fundamental features, with the second introducing additional functionalities that have already been the subject of critical discussion in current literature. A differentiation by known groups can be performed on the assumption that the second prototype will cause significantly more concern among the participants. This approach was first proposed by Churchill et al. [11] and was analogously in previous scale development processes [55, 78]. The results of the second survey, divided into the two prototypes and analyzed separately, prove this assumption to be correct. After assessing that a normal distribution could be assumed with a Shapiro-Wilk test (W = 0.99, $p = 0.34$) and that homogeneity of variances is given with Levene's test (L$(1, 96) = 1.38, p = 0.24$), an independent t-test ($t(96) = -3.36, p = 0.001$) revealed that the resulting score of the MRC Questionnaire for the first scenario ($\bar{x}_{\text{MRC}} = 29.1, s_{\text{MRC}} = 6.96$) was significantly lower than for the second scenario ($\bar{x}_{\text{MRC}} = 33.6, s_{\text{MRC}} = 6.3$). Table 3 shows the full results of this step.

5.3.2 Convergent/Divergent Validity. To compare the results from the MRC Questionnaire with established questionnaires, participants evaluated the presented hypothetical prototypes using not only the MRC but also the PCTS [78] and the UEQ [66].

As the PCTS is one of the only scales that explicitly sets out to measure negative sentiments towards technologies, a high correlation between the MRC Questionnaire and it is desired. The PCTS assesses the perceived creepiness of a technology in regards to the three factors *Implied Malice*, *Undesirability*, and *Unpredicability*. One might assume that when individuals perceive a technology as having potential security or privacy vulnerabilities, they may consider it undesirable. The presence of security and privacy concerns might undermine the technology's trustworthiness, potentially making it less predictable in turn. Furthermore, when users perceive a technology as having social implications that may disrupt or harm societal norms, they may interpret these consequences as indicative of implied malice. To the best of the authors' knowledge, there is currently no other questionnaire specifically designed to evaluate negative sentiments toward emerging technologies directly. As Figure 3 shows, the MRC and PCTS correlate ($r = 0.58, 95\%$ CI $= 0.43, 0.70$), indicating that the perceived feeling of creepiness evoked by an MR system and the magnitude of concerns raised in relation to it are both impacted similarly. While a simple correlation test cannot prove the above-mentioned hypothesized causations, the scales do correlate as expected. While we assume that the PCTS assesses the feelings (i.e., invoked creepiness) that are a reaction to the system's concerns, and with this its inherent properties, further research is needed to prove this connection.

We incorporated the UEQ [66] for another comparative assessment. The comparison with the UEQ is particularly valuable due to its widespread use and established reputation as a comprehensive tool for assessing overall user experience, encompassing classical usability aspects as well as user experience dimensions. Among the available questionnaires, the UEQ was chosen for its versatility and applicability across various technological contexts, providing a well-established benchmark against which the effectiveness and
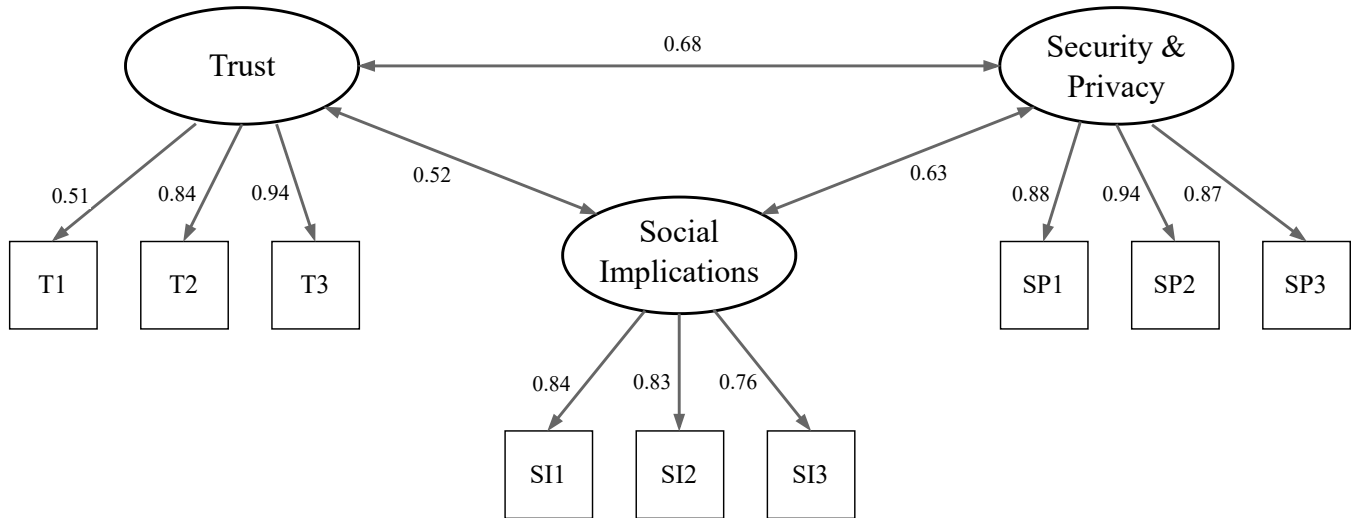
Figure 2: The result of the CFA confirms this three-factor model for the scale, with moderate correlations between the subscales and mostly high item coefficients.

Table 3: Differentiation by known groups.

Scale/Subscale	Scenario 1		Scenario 2		Independent t-Test
	$\bar{x}$	s	$\bar{x}$	s	
MRC	29.1	6.96	33.6	6.3	$t(96) = -3.36, p = 0.001$
Security & Privacy	9.86	3.2	11.7	2.63	$t(97) = -3.17, p = 0.002$
Social Implications	9.76	3.22	11.9	2.7	$t(97) = -3.53, p < 0.001$
Trust	9.49	2.19	10	2.45	$t(97) = -1.09, p = 0.278$

specificity of the MRC questionnaire can be meaningfully evaluated. While factors like efficiency or perspicuity can be hard to assess through a text description and a mockup image only, we specifically focused on the two hedonic qualities, those being stimulation and novelty. Our interest in these hedonic qualities arises from the hypothesis that when a new device is perceived as subpar or unneeded, users may harbor more concerns. Conversely, when a new system is viewed as exceedingly novel and futuristic, concerns may stem more from unfamiliarity than from actual substantive concerns regarding the device. However, the test results reveal that both stimulation ($r = 0.22, 95\%$ CI $= 0.02, 0.39$) and novelty ($r = 0.17, 95\%$ CI $= -0.03, 0.35$) exhibit a low correlation with the MRC, suggesting that concerns related to MR systems encompass more than just stimulation and novelty. For completeness sake, all UEQ scales are shown in Figure 3.

5.4 Third Survey

In addition to the Cronbach's alphas reported in Section 5.2 as tests of reliability, we opted for performing one further test-retest reliability evaluation by conducting a final third survey.

5.4.1 Participants. Instead of using Prolific for recruitment, as in the first two surveys, participants were invited to take part through institute mailing lists and snowball sampling. In the end, a total

of $n = 12$ people participated in the online survey, which took approximately 5 minutes to complete. Again, the participation was entirely voluntary, and the option to withdraw from the survey was available throughout. No compensation was given for the third survey. The average age of participants was roughly 27 years ($\bar{x} = 27.25, s = 4.0$), with two-thirds ($n = 8$) identifying as female and the rest identifying as male and all currently residing in countries of the European Union. As noted by Mejia and Yarosh [55], while it often poses difficulty to recruit enough people for two survey runs, and this usually being the reason why a test-retest evaluation is omitted, we too opted for still performing this validation, even if only a smaller sample size could be achieved. The time between the two runs was set to be at least ten days to ensure a long enough time between the two reflections on the presented prototype.

5.4.2 Survey Structure. Participants were shown one hypothetical prototype, for which, based on the explained feature set, relatively high values were to be expected. It consisted of an AR social application that enabled users to receive automatic information about their conversation partners through facial recognition. Additionally, it provided the functionality to rate individuals and conversations
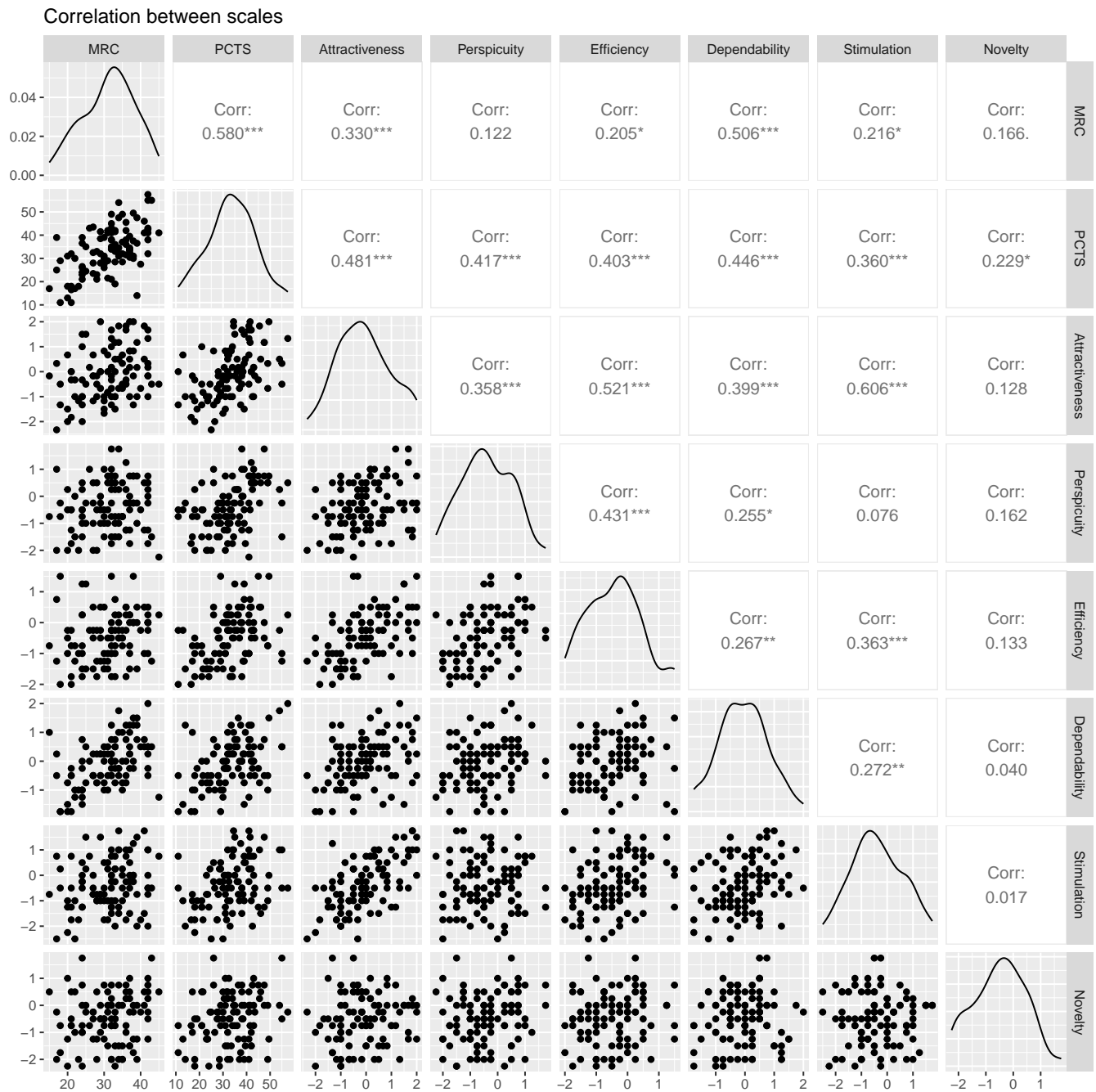
Correlation between scales

Figure 3: The main diagonal shows the histograms of each metric; the lower triangular shows the correlation plots between the metrics, and the upper triangular shows the corresponding r-values for the different scales under comparison. It is evident that the MRC and PCTS [78] highly correlate. Furthermore, the MRC correlates with both the *Attractiveness* and *Dependability* subscales of the UEQ [66].

publicly. This concept was based on related work [32, 37] and a now-defunct social media platform with a similar set of features[8].

---
[8]https://techcrunch.com/2016/03/08/controversial-people-rating-app-peeple-goes-live-has-a-plan-to-profit-from-users-negative-reviews, last accessed on 2023-12-12.

The MR system was described in a 260-word description, and a mockup of a potential interface for such an application was supplied. Afterward, participants were instructed to state their agreement

with each of the items of the final MRC Questionnaire as shown in Table 2.

5.5 Test-Retest Reliability

As suggested by Rousson et al. [64], we evaluated the Pearson product-moment correlations for both the subscales and the MRC Questionnaire as a whole. While the *Security & Privacy* only showed an acceptable correlation for a test-retest context [12], the two other subscales showed much higher correlations. In total, the MRC Questionnaire exhibits a moderate to excellent test-retest reliability ($r = 0.85$, 95% CI = 0.54, 0.96). The correlation plots and respective correlation values are shown in Figure 4. Based on this reliability test, especially considering the small sample size, it can be assumed that the MRC Questionnaire shows temporal stability and can be used in repeated-measures studies.

6 DISCUSSION

In this section, we present instructions on using the MRC Questionnaire and interpreting its results. Furthermore, we explain the limitations of our approach and the scale as well as ideas for future enhancements.

6.1 Scoring

The MRC Questionnaire is scored on a 5-point Likert scale, ranging from *Strongly disagree* (1) to *Strongly agree* (5). All items of the *Trust* subscale are reverse-coded.

$$\text{MRC} = \text{MRC}_{\text{SP}} + \text{MRC}_{\text{SI}} + \text{MRC}_{\text{T}}$$
$$\text{with } \text{MRC}_{\text{SP}} = \text{SP1} + \text{SP2} + \text{SP3}$$
$$\text{and } \text{MRC}_{\text{SI}} = \text{SI1} + \text{SI2} + \text{SI3}$$
$$\text{and } \text{MRC}_{\text{T}} = \text{T1}_R + \text{T2}_R + \text{T3}_R$$

As a result, the scale's range spans from 9 as the lowest score to 45 as the highest. Elevated scores signify higher concerns associated with the MR system.

6.2 Guidelines and Limitations to Administering the Scale

A measuring instrument, such as the presented MRC Questionnaire, which is designed to assess concerns related to MR systems, can be immensely valuable for the research, development, and improvement of these technologies. Such an instrument might serve as a crucial tool in several ways:

This scale is intentionally designed not to assess the specific, objective problems or risks associated with a technology but rather to focus on user apprehensions and concerns. Its primary purpose is to measure the subjective perceptions and feelings of users regarding a technology, particularly any unease or worries they may experience. By concentrating on user apprehensions, the scale aims to capture the emotional and psychological aspects of how MR systems might be perceived even before actual user experiences can be gathered. It recognizes that people's perceptions and concerns can vary widely, even when faced with similar objective risks or

issues. Therefore, the scale provides a means to gauge how users interpret and respond to these risks on a personal level.

Conversely, it can also be used to assess actual implementations. Users' apprehensions often reveal pain points or areas of discomfort about the technology at hand. This information is valuable for pinpointing specific issues that may need addressing, whether they relate to security, privacy, social implications, or the inherent trust in the system. User concerns can also guide the development of educational materials or resources to help users understand the technology better. Addressing misconceptions or alleviating fears through education can contribute to a more positive user experience. In summary, while the scale's primary focus is on assessing user apprehensions and perceptions, it can serve as a versatile tool for evaluating new parts of the user experience in actual technology implementations, which other scales currently do not assess. By understanding and addressing user concerns, developers can enhance the overall quality and acceptance of MR systems and other technologies.

The preceding evaluation suggests that applying the MRC Questionnaire is suitable for both between-subject and within-subject studies, as well as for repeated-measures studies. Although the analysis of the subscales generally presents favorable results for evaluating them on their own, we do not explicitly recommend this application. The intentional brevity of the scale serves the purpose of offering a quick initial insight into potential user concerns. However, the precise nature of these concerns should be explored through additional qualitative research and is likely to be highly specific to the particular MR system under consideration. As illustrated by the conceptual model in Section 3, the realm of potential reasons for concern is too expansive to encompass within a single scale suitable for a wide range of applications. Once again, this scale is designed primarily to provide an initial understanding of potential user concerns.

Finally, it is crucial to emphasize that this scale is not inherently linked to the acceptability of a system. Although we assume that the absence of concerns can certainly impact acceptability, numerous other factors may come into play. For this, other scales and questionnaires, like the ones presented in Section 2.3 and Section 5.3.2, should be used in conjunction with the MRC Questionnaire.

6.3 Limitations of the Development Process

Next to the aforementioned limitations to how the scale can be used and evaluated, we acknowledge that the development process of the MRC Questionnaire may be subject to certain limitations, too. First and foremost, the exploratory factor analysis, as well as all subsequent evaluation stages, was conducted during a period when MR technologies were gradually making their way toward broader public acceptance. The trajectory of development and widespread adoption of these devices in the coming years remains uncertain. Consequently, it is likely that opinions, perceptions, and concerns will change over time. Therefore, a reevaluation of the scale may become necessary in the future.

Much like the PCTS [78], we opted to concentrate on developing a scale that evaluates users' concerns and apprehensions immediately after the first introduction to that MR system. Due to this, the suitability of the MRC for long-term studies remains uncertain.
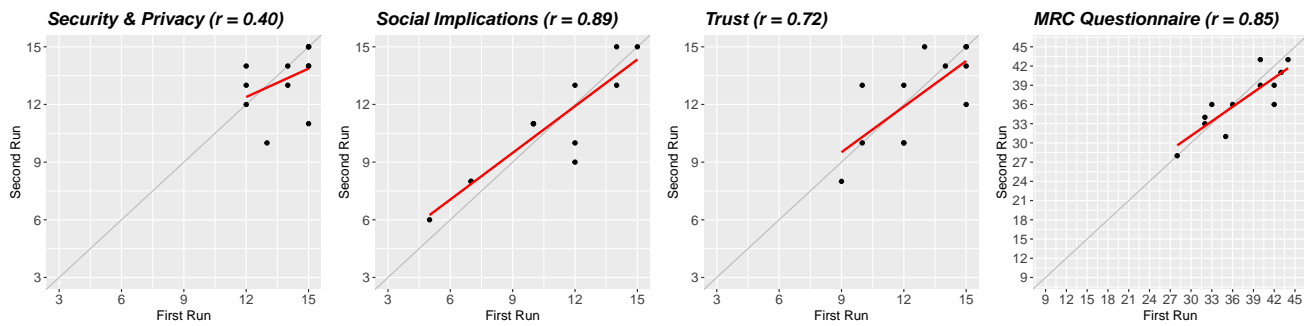
Figure 4: The different subscale and overall scores for both runs of the third survey. Furthermore, the Pearson product-moment correlation is given for all plots.

While we expect that the scale might have the potential to measure how user concerns change over time, it is essential to note that this capability cannot be definitively affirmed at the time being.

Additionally, the study primarily involved participants from countries with a Western cultural background, and as the surveys were conducted online, all participants possessed at least a basic understanding of current consumer electronics. While we hope for the scale to have relevance in diverse cultural contexts and among individuals with varying levels of familiarity with consumer electronics, we cannot guarantee this outcome. Ideally, future research will address this issue, facilitating cross-cultural and demographic comparisons of different concerns and apprehensions that people might have regarding MR systems.

The lack of real exposure testing introduces uncertainty of external factors (e.g., user context [65] or situatively perceived cognitive workload [48] during MR use), regarding the questionnaire's performance for capturing concerns when interacting with MR systems. The potential biases or deviations in user responses under actual MR exposure conditions raise consideration since they could impact the questionnaire's reliability and validity in such contexts (cf. [4, 46, 74] for biased study data when users have specific expectations towards novel technologies). To address this limitation, future research should prioritize conducting evaluations with participants exposed to operational MR systems using the MRC questionnaire. This approach will provide a more comprehensive understanding of the MRC questionnaire's effectiveness in capturing user experiences. Additionally, incorporating user feedback from authentic MR interactions will contribute to refining the questionnaire for increased applicability and relevance in practical settings.

7 CONCLUSION

We present a measurement tool designed to evaluate user concerns and apprehensions regarding MR systems. Initially, we constructed a conceptual model outlining potential concerns associated with MR systems, drawing insights from existing research. Subsequently, we engaged in two rounds of expert feedback to generate a comprehensive set of survey items. A total of three surveys were conducted to first reduce this set of items and then evaluate the final MRC Questionnaire.

The questionnaire shows high internal consistency, adequate temporal stability, and high convergent and divergent validity. It

serves as a valuable instrument for assessing the initial concerns individuals may harbor when encountering a new MR system. Furthermore, its intentional brevity enables its application in various studies and situations where an initial understanding of potential apprehensions is required.

We aspire for this scale to help researchers and developers cultivate a constructive approach to these concerns. It can serve as a tool to ensure that new MR artifacts and applications transparently convey their intentions, features, and potential impact on both users and bystanders. While this assessment could prove beneficial for educational purposes, it is essential to emphasize that addressing potential concerns primarily falls within the realm of technological development rather than solely relying on user adaptation or adjustment.

The questionnaire and supplementary material are openly accessible on the research group's website[9].

ACKNOWLEDGMENTS

REFERENCES
[1] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2016. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1968–1979. https://doi.org/10.1145/2858036.2858226

[2] Gaurav Bhorkar. 2017. A Survey of Augmented Reality Navigation. https://doi.org/10.48550/arXiv.1708.05006 arXiv:1708.05006 [cs]

[3] Godfred O. Boateng, Torsten B. Neilands, Edward A. Frongillo, Hugo R. Melgar-Quiñonez, and Sera L. Young. 2018. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health* 6 (2018), 149.

[4] Walter Boot, Daniel Simons, Cary Stothart, and Cassie Stutts Berry. 2013. The Pervasive Problem With Placebos in Psychology Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects. *Perspectives on Psychological Science* 8 (07 2013), 445–454. https://doi.org/10.1177/1745691613491271

[5] Ulla Bunz, Jonmichael Seibert, and Joshua Hendrickse. 2021. From TAM to AVRTS: Development and Validation of the Attitudes toward Virtual Reality Technology Scale. *Virtual Reality* 25, 1 (March 2021), 31–41. https://doi.org/10.1007/s10055-020-00437-7

[6] Sebastian Büttner, Henrik Mucha, Markus Funk, Thomas Kosch, Mario Aehnelt, Sebastian Robert, and Carsten Röcker. 2017. The Design Space of Augmented and Virtual Reality Applications for Assistive Environments in Manufacturing: A Visual Approach. In *Proceedings of the 10th International Conference on PErvasive*

---

[9]https://hcistudio.org/mrc-questionnaire

Technologies Related to Assistive Environments (Island of Rhodes, Greece) *(PETRA '17)*. Association for Computing Machinery, New York, NY, USA, 433–440. https://doi.org/10.1145/3056540.3076193

[7] Li Cai, Seung Won Chung, and Taehun Lee. 2023. Incremental Model Fit Assessment in the Case of Categorical Data: Tucker–Lewis Index for Item Response Theory Modeling. *Prevention Science* 24, 3 (April 2023), 455–466. https://doi.org/10.1007/s11121-021-01253-4

[8] Peter Casey, Ibrahim Baggili, and Ananya Yarramreddy. 2021. Immersive Virtual Reality Attacks and the Human Joystick. *IEEE Transactions on Dependable and Secure Computing* 18, 2 (March 2021), 550–562. https://doi.org/10.1109/TDSC.2019.2907942

[9] Raymond B. Cattell. 1966. The Scree Test For The Number Of Factors. *Multivariate Behavioral Research* 1, 2 (April 1966), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

[10] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun. 2011. Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 41, 6 (Nov. 2011), 1052–1063. https://doi.org/10.1109/TSMCA.2011.2116000

[11] Gilbert A. Churchill. 1979. A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research* 16, 1 (1979), 64–73. https://doi.org/10.2307/3150876 jstor:3150876

[12] Jacob Cohen. 2009. *Statistical Power Analysis for the Behavioral Sciences* (2. ed., reprint ed.). Psychology Press, New York, NY.

[13] Andrew L. Comrey. 1988. Factor-Analytic Methods of Scale Development in Personality and Clinical Psychology. *Journal of Consulting and Clinical Psychology* 56, 5 (1988), 754–761. https://doi.org/10.1037/0022-006X.56.5.754

[14] Matthew Corbett, Brendan David-John, Jiacheng Shang, Y. Charlie Hu, and Bo Ji. 2024. Securing Bystander Privacy in Mixed Reality While Protecting the User Experience. *IEEE Security & Privacy* 22, 1 (2024), 33–42. https://doi.org/10.1109/MSEC.2023.3331649

[15] Jose M. Cortina. 1993. What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology* 78, 1 (1993), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

[16] Areti Damala and Nenad Stojanovic. 2012. Tailoring the Adaptive Augmented Reality (A2R) Museum Visit: Identifying Cultural Heritage Professionals' Motivations and Needs. In *Nternational Symposium on Mixed and Augmented Reality 2012 (ISMAR 2012)*. ISMAR-AMH, Atlanta, USA, 71–80.

[17] Fred D. Davis. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 3 (1989), 319–340. https://doi.org/10.2307/249008 jstor:249008

[18] Fred D. Davis, Richard P. Bagozzi, and Paul R. Warshaw. 1989. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science* 35, 8 (1989), 982–1003. jstor:2632151

[19] Jaybie A. De Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. 2019. Security and Privacy Approaches in Mixed Reality: A Literature Survey. *Comput. Surveys* 52, 6 (Oct. 2019), 110:1–110:37. https://doi.org/10.1145/3359626

[20] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. 2011. A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfillment of Privacy Requirements. *Requirements Engineering* 16, 1 (March 2011), 3–32. https://doi.org/10.1007/s00766-010-0115-7

[21] Chloe Eghtebas, Gudrun Klinker, Susanne Boll, and Marion Koelle. 2023. Co-Speculating on Dark Scenarios and Unintended Consequences of a Ubiquitous(Ly) Augmented Reality. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2392–2407. https://doi.org/10.1145/3563657.3596073

[22] João Marcelo Evangelista Belo, Anna Maria Feit, Tiare Feuchtner, and Kaj Grønbæk. 2021. XRgonomics: Facilitating the Creation of Ergonomic 3D Interfaces. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 290, 11 pages. https://doi.org/10.1145/3411764.3445349

[23] Sebastian Feger, Lars Semmler, Albrecht Schmidt, and Thomas Kosch. 2022. ElectronicsAR: Design and Evaluation of a Mobile and Tangible High-Fidelity Augmented Electronics Toolkit. *Proc. ACM Hum.-Comput. Interact.* 6, ISS, Article 587 (nov 2022), 22 pages. https://doi.org/10.1145/3567740

[24] Archie A. George, Gene E. Hall, and Suzanne Stiegelbauer. 2008. *Measuring Implementation in Schools: The Stages of Concern Questionnaire* (2. print. with minor additions and corr ed.). Southwest Educational Development Laboratory, Austin, Tex.

[25] Yiannis Georgiou and Eleni A. Kyza. 2017. The Development and Validation of the ARI Questionnaire. *International Journal of Human-Computer Studies* 98, C (Feb. 2017), 24–37. https://doi.org/10.1016/j.ijhcs.2016.09.014

[26] Rebecca A. Grier, Aaron Bangor, Philip Kortum, and S. Camille Peres. 2013. The System Usability Scale: Beyond Standard Usability Testing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57, 1 (Sept. 2013), 187–191. https://doi.org/10.1177/1541931213571042

[27] Jan Gugenheimer, Wen-Jie Tseng, Abraham Hani Mhaidli, Jan Ole Rixen, Mark McGill, Michael Nebeling, Mohamed Khamis, Florian Schaub, and Sanchari Das. 2022. Novel Challenges of Safety, Security and Privacy in Extended Reality. In

Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3491101.3503741

[28] Yuntao Guo, Shubham Agrawal, Srinivas Peeta, and Irina Benedyk. 2021. Safety and Health Perceptions of Location-Based Augmented Reality Gaming App and Their Implications. *Accident Analysis & Prevention* 161 (Oct. 2021), 106354. https://doi.org/10.1016/j.aap.2021.106354

[29] Nidhi Gupta, Arnout R.H. Fischer, and Lynn J. Frewer. 2012. Socio-Psychological Determinants of Public Acceptance of Technologies: A Review. *Public Understanding of Science* 21, 7 (Oct. 2012), 782–795. https://doi.org/10.1177/0963662510392485

[30] David Harborth and Sebastian Pape. 2021. Investigating Privacy Concerns Related to Mobile Augmented Reality Apps – A Vignette Based Online Experiment. *Computers in Human Behavior* 122 (Sept. 2021), 106833. https://doi.org/10.1016/j.chb.2021.106833

[31] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003: Interaktion in Bewegung*, Gerd Szwillus and Jürgen Ziegler (Eds.). Vieweg+Teubner Verlag, Wiesbaden, 187–196. https://doi.org/10.1007/978-3-322-80058-9_19

[32] Mohammed E. Hoque. 2012. My Automated Conversation Helper (MACH): Helping People Improve Social Skills.. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*. ACM, Santa Monica California USA, 313–316. https://doi.org/10.1145/2388676.2388745

[33] John L. Horn. 1965. A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika* 30, 2 (June 1965), 179–185. https://doi.org/10.1007/BF02289447

[34] Michael Howard and Steve Lipner. 2006. *The Security Development Lifecycle: SDL, a Process for Developing Demonstrably More Secure Software*. Microsoft Press, Redmond, WA.

[35] C.E. Hughes, C.B. Stapleton, D.E. Hughes, and E.M. Smith. 2005. Mixed Reality in Education, Entertainment, and Training. *IEEE Computer Graphics and Applications* 25, 6 (Nov. 2005), 24–30. https://doi.org/10.1109/MCG.2005.139

[36] Muhammad Hussain, Jaehyun Park, and Hyun K. Kim. 2023. Augmented Reality Sickness Questionnaire (ARSQ): A Refined Questionnaire for Augmented Reality Environment. *International Journal of Industrial Ergonomics* 97 (Sept. 2023), 103495. https://doi.org/10.1016/j.ergon.2023.103495

[37] Katherine Isbister, Hideyuki Nakanishi, Toru Ishida, and Cliff Nass. 2000. Helper Agent: Designing an Assistant for Human-Human Interaction in a Virtual Meeting Space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. Association for Computing Machinery, New York, NY, USA, 57–64. https://doi.org/10.1145/332040.332407

[38] Jiun-Yin Jian, Ann Bisantz, and Colin Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics* 4 (March 2000), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

[39] Lena Jingen Liang and Statia Elliot. 2021. A Systematic Review of Augmented Reality Tourism Research: What Is Now and What Is Next? *Tourism and Hospitality Research* 21, 1 (Jan. 2021), 15–30. https://doi.org/10.1177/1467358420941913

[40] Christos Kalloniatis, Evangelia Kavakli, and Stefanos Gritzalis. 2008. Addressing Privacy Requirements in System Design: The PriS Method. *Requirements Engineering* 13, 3 (Sept. 2008), 241–255. https://doi.org/10.1007/s00766-008-0067-3

[41] Christopher Katins, Sebastian S. Feger, and Thomas Kosch. 2023. Exploring Mixed Reality in General Aviation to Support Pilot Workload. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 116, 7 pages. https://doi.org/10.1145/3544549.3585742

[42] Christopher Katins, Sebastian Stefan Feger, and Thomas Kosch. 2023. Pilots' Considerations Regarding Current Generation Mixed Reality Headset Use in General Aviation Cockpits. In *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia* (Vienna, Austria) *(MUM '23)*. Association for Computing Machinery, New York, NY, USA, 159–165. https://doi.org/10.1145/3626705.3627785

[43] Mara Kaufeld, Martin Mundt, Sarah Forst, and Heiko Hecht. 2022. Optical See-through Augmented Reality Can Induce Severe Motion Sickness. *Displays* 74 (Sept. 2022), 102283. https://doi.org/10.1016/j.displa.2022.102283

[44] Hyun K. Kim, Jaehyun Park, Yeongcheol Choi, and Mungyeong Choe. 2018. Virtual Reality Sickness Questionnaire (VRSQ): Motion Sickness Measurement Index in a Virtual Reality Environment. *Applied Ergonomics* 69 (May 2018), 66–73. https://doi.org/10.1016/j.apergo.2017.12.016

[45] Kiyoshi Kiyokawa. 2007. An Introduction to Head Mounted Displays for Augmented Reality. In *Emerging Technologies of Augmented Reality: Interfaces and Design*. IGI Global, Osaka University, Japan, 43–63. https://doi.org/10.4018/978-1-59904-066-0.ch003

[46] Agnes M. Kloft, Robin Welsch, Thomas Kosch, and Steeven Villa. 2023. "AI enhances our performance, I have no doubt this one will do the same": The Placebo effect is robust to negative descriptions of AI. arXiv:2309.16606 [cs.HC]

[47] Pascal Knierim, Albrecht Schmidt, and Thomas Kosch. 2020. Demonstrating Thermal Flux: Using Mixed Reality to Extend Human Sight by Thermal Vision. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia* (Duisburg-Essen, Germany) *(MUM '20)*. ACM, New York, NY, USA, 348–350. https://doi.org/10.1145/3428361.3431196

[48] Thomas Kosch, Jakob Karolus, Johannes Zagermann, Harald Reiterer, Albrecht Schmidt, and Paweł W. Woźniak. 2023. A Survey on Measuring Cognitive Workload in Human-Computer Interaction. *Comput. Surveys* 55, 13s (July 2023), 283:1–283:39. https://doi.org/10.1145/3582272

[49] Thomas Kosch, Andrii Matviienko, Florian Müller, Jessica Bersch, Christopher Katins, Dominik Schön, and Max Mühlhäuser. 2022. NotiBike: Assessing Target Selection Techniques for Cyclist Notifications in Augmented Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (2022), 1–24.

[50] Chi-Jung Lee and Hung-Kuo Chu. 2018. Dual-MR: Interaction with Mixed Reality Using Smartphones. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3281505.3281618

[51] Shu-Sheng Liaw and Hsiu-Mei Huang. 2003. An Investigation of User Attitudes toward Search Engines as an Information Retrieval Tool. *Computers in Human Behavior* 19, 6 (Nov. 2003), 751–765. https://doi.org/10.1016/S0747-5632(03)00009-8

[52] D. Betsy McCoach, Robert K. Gable, and John P. Madura. 2013. *Instrument Development in the Affective Domain: School and Corporate Applications*. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-7135-6

[53] Graeme McLean and Mohammed Aldossary. 2023. Digital Tourism Consumption: The Role of Virtual Reality (VR) Vacations on Consumers' Psychological Wellbeing: An Abstract. In *Optimistic Marketing in Challenging Times: Serving Ever-Shifting Customer Needs (Developments in Marketing Science: Proceedings of the Academy of Marketing Science)*, Bruna Jochims and Juliann Allen (Eds.). Springer Nature Switzerland, Cham, 143–144. https://doi.org/10.1007/978-3-031-24687-6_53

[54] Arian Mehrfard, Javad Fotouhi, Giacomo Taylor, Tess Forster, Nassir Navab, and Bernhard Fuerst. 2019. A Comparative Analysis of Virtual Reality Head-Mounted Display Systems. arXiv:1912.02913 [cs]

[55] Kenya Mejia and Svetlana Yarosh. 2017. A Nine-Item Questionnaire for Measuring the Social Disfordance of Mediated Social Touch Technologies. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 77:1–77:17. https://doi.org/10.1145/3134712

[56] Thomas Moser, Markus Hohlagschwandtner, Gerhard Kormann-Hainzl, Sabine Pölzlbauer, and Josef Wolfartsberger. 2019. Mixed Reality Applications in Industry: Challenges and Research Areas. In *Software Quality: The Complexity and Challenges of Software Engineering and Software Quality in the Cloud (Lecture Notes in Business Information Processing)*, Dietmar Winkler, Stefan Biffl, and Johannes Bergsmann (Eds.). Springer International Publishing, Cham, 95–105. https://doi.org/10.1007/978-3-030-05767-1_7

[57] Wolfgang Narzt, Gustav Pomberger, Alois Ferscha, Dieter Kolb, Reiner Müller, Jan Wieghardt, Horst Hörtner, and Christopher Lindinger. 2006. Augmented Reality Navigation Systems. *Universal Access in the Information Society* 4, 3 (March 2006), 177–187. https://doi.org/10.1007/s10209-005-0017-5

[58] Adam Nowak, Pascal Knierim, Andrzej Romanowski, Albrecht Schmidt, and Thomas Kosch. 2020. What does the Oscilloscope Say?: Comparing the Efficiency of In-Situ Visualisations during Circuit Analysis. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. ACM, New York, NY, USA, 1–7. https://doi.org/10.1145/3334480.3382890

[59] Yusuke Numazaki, See-Sheng Toh, and Masanobu Endoh. Decmber 8-10, 2017. VR Entertainment System "Ideal Vacation": A Game Designing Focused on the Sense of Presence. In *The 2nd International Conference on Culture Technology (ICCT)*. International Association for Convergence Science & Technology, Tokyo, Japan, 163–166.

[60] S. Ram and Jagdish N. Sheth. 1989. Consumer Resistance to Innovations: The Marketing Problem and Its Solutions. *Journal of Consumer Marketing* 6, 2 (Jan. 1989), 5–14. https://doi.org/10.1108/EUM0000000002542

[61] Tenko Raykov and George A. Marcoulides. 2011. Introduction to Psychometric Theory. https://www.routledge.com/Introduction-to-Psychometric-Theory/Raykov-Marcoulides/p/book/9780415878227.

[62] Holger Regenbrecht and Thomas Schubert. 2021. Measuring Presence in Augmented Reality Environments: Design and a First Test of a Questionnaire. https://doi.org/10.48550/arXiv.2103.02831 arXiv:2103.02831 [cs]

[63] Alexandra Rese, Daniel Baier, Andreas Geyer-Schulz, and Stefanie Schreiber. 2017. How Augmented Reality Apps Are Accepted by Consumers: A Comparative Analysis Using Scales and Opinions. *Technological Forecasting and Social Change* 124 (Nov. 2017), 306–319. https://doi.org/10.1016/j.techfore.2016.10.010

[64] Valentin Rousson, Theo Gasser, and Burkhardt Seifert. 2002. Assessing Intrarater, Interrater and Test–Retest Reliability of Continuous Measurements. *Statistics in Medicine* 21, 22 (2002), 3431–3446. https://doi.org/10.1002/sim.1253

[65] B. Schilit, N. Adams, and R. Want. 1994. Context-Aware Computing Applications. In *1994 First Workshop on Mobile Computing Systems and Applications*. IEEE, Santa Cruz, CA, USA, 85–90. https://doi.org/10.1109/WMCSA.1994.16

[66] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40. https://doi.org/10.9781/ijimai.2017.445

[67] Dominik Schön, Thomas Kosch, Florian Müller, Martin Schmitz, Sebastian Günther, Lukas Bommhardt, and Max Mühlhäuser. 2023. Tailor Twist: Assessing Rotational Mid-Air Interactions for Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3544548.3581461

[68] Noora Shrestha. 2021. Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics* 9, 1 (Jan. 2021), 4–11. https://doi.org/10.12691/ajams-9-1-2

[69] Mel Slater, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, Graham Breen, Raz Schwarz, William Steptoe, Dalila Szostak, Shivashankar Halan, Deborah Fox, and Jeremy Silver. 2020. The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality* 1 (March 2020), 1. https://doi.org/10.3389/frvir.2020.00001

[70] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. 2019. What Is Mixed Reality?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, Glasgow, Scotland Uk, 1–15. https://doi.org/10.1145/3290605.3300767

[71] Derianna Thomas and Lars Erik Holmquist. 2021. Is Functionality All That Matters? Examining Everyday User Opinions of Augmented Reality Devices. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, Lisbon, Portugal, 232–237. https://doi.org/10.1109/VRW52623.2021.00050

[72] Viswanath Venkatesh and Hillol Bala. 2008. Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences* 39, 2 (2008), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

[73] Viswanath Venkatesh and Fred D. Davis. 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies | Management Science. https://pubsonline.informs.org/doi/abs/10.1287/mnsc.46.2.186.11926.

[74] Steeven Villa, Thomas Kosch, Felix Grelka, Albrecht Schmidt, and Robin Welsch. 2023. The Placebo Effect of Human Augmentation: Anticipating Cognitive Augmentation Increases Risk-Taking Behavior. *Computers in Human Behavior* 146, C (2023), 107787. https://doi.org/10.1016/j.chb.2023.107787

[75] Steeven Villa, Jasmin Niess, Albrecht Schmidt, and Robin Welsch. 2023. Society's Attitudes Towards Human Augmentation and Performance Enhancement Technologies (SHAPE) Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–23.

[76] Graham Wilson and Mark McGill. 2018. Violent Video Games in Virtual Reality: Re-Evaluating the Impact and Rating of Interactive Experiences. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 535–548. https://doi.org/10.1145/3242671.3242684

[77] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (June 1998), 225–240. https://doi.org/10.1162/105474698565686

[78] Paweł W. Woźniak, Jakob Karolus, Florian Lang, Caroline Eckerth, Johannes Schöning, Yvonne Rogers, and Jasmin Niess. 2021. Creepy Technology:What Is It and How Do You Measure It?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–13. https://doi.org/10.1145/3411764.3445299

[79] Guangjian Zhang and Kristopher J. Preacher. 2015. Factor Rotation and Standard Errors in Exploratory Factor Analysis - Guangjian Zhang, Kristopher J. Preacher, 2015. https://journals.sagepub.com/doi/10.3102/1076998615606098.