2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) | 979-8-3315-0691-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISMAR-Adjunct64951.2024.00020

Enhancing Replicability in XR HCI Studies: A Survey-Based Approach

Zahra Borhani* Francisco R. Ortega†

Colorado State University

Abstract

To ensure the reliability and validity of research findings, the results of studies should be replicable. However, the fields of HCI and Extended Reality (XR) research currently lack coherent strategies to enhance replicability. Although some studies have explored the challenges of replication in HCI and Virtual Reality (VR) research, no systematic review has specifically addressed replication issues in XR applications. This paper presents a systematic literature review encompassing 15 studies, to identify when replication is necessary in XR HCI research and explore methods to enhance replicability.

Index Terms: Replication, Human-Computer Interaction (HCI), Extended Reality (XR), Virtual Reality (VR), Augmented Reality (AR), Systematic Review

1 Introduction

Replication is a cornerstone of scientific research, ensuring the reliability and validity of findings. This paper distinguishes between three types of replication: direct, conceptual, and constructive replication. Direct replication seeks to reproduce the methods and conditions of the original study as precisely as possible to verify if the same results can be achieved [7]. Conceptual replication employs different operationalizations, variables, experimental designs, and participants to test the theory proposed in the original study [7]. While a conceptual replication aims at testing the same theoretical ideas using different methods, a Constructive replication is about improving or refining the original study's methods or findings, often by modifying the experimental design to address weaknesses of the original study or explore new dimensions [1].

Scientific research should be reproducible [25], however, HCI and Extended Reality (XR) research lacks coherent approaches to increase replicability. Rzig et al. investigated 314 open-source VR applications for their replicability. They found that 79% of the VR projects had no automatic tests, and those that did had lower median functional-method to test-method ratios compared to other project types [27].

Although some previous studies investigated the replication challenges in HCI and VR research [27, 14, 31], there is currently no survey reviewing the replication problem through a systematic review for XR applications. Rzig et al. conducted an empirical study on 314 open-source VR applications to investigate the challenges of automated testing in the wild [27]. Hepperle et al. explored possible solutions to reduce the human factor in VR research [14]. In this paper, we summarized the need for the replication of a study and methods to enhance the replication process. This work aims to answer the following research questions (RQs) through a systematic literature review of studies exploring replication in XR systems:

- **RQ1:** What are the key motivations for replicating an XR HCI research study?
- **RQ2:** What strategies can be implemented to improve the replicability of an XR HCI study?

*e-mail: zahra.borhani@colostate.edu
†e-mail:F.Ortega@colostate.edu

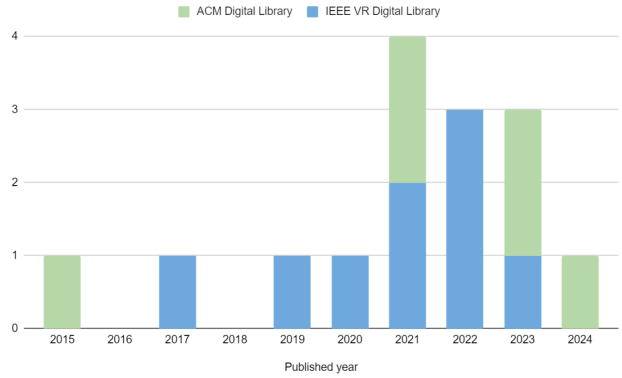

Figure 1: Distribution of publications

2 Method

This paper reviewed 15 studies related to replication in XR, selected through our systematic investigation methods. Each chosen study either replicated a previous experiment or provided materials to facilitate replication. Our literature corpus was sourced from the ACM Digital Library and IEEE Xplore search engines, with no restrictions on publication year. These are two leading publishers in the augmented reality domain[8]. Figure 1 shows the distribution of these 15 surveyed papers over previous years.

We employed a filtering method similar to [4]. The initial selection criterion required the paper's abstract to include at least one of our specified search queries. Keywords included "augmented/virtual/extended/mixed reality" and "replication." In the second step, we excluded research papers not published in XR or Human-Computer Interaction (HCI) venues. We investigated venues relevant to XR HCI from the IEEE and ACM Digital Libraries, encompassing the IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE Transactions on Haptics, IEEE Virtual Reality (VR), IEEE Transactions on Visualization and Computer Graphics (TVCG), IEEE Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), IEEE Communications Surveys and Tutorials, ACM Conference on Human Factors in Computing Systems (CHI), ACM Conference on Computer Supported Cooperative Work (CSCW), and ACM Symposium on User Interface Software and Technology (UIST). In the third phase, we reviewed the entire text of each paper to ensure they met our criteria. In the final step, we discarded the duplicate papers, which resulted in 15 papers in the final phase. Figure 2 illustrates our literature review process.

Out of 15 papers, nine were selected for replicating a previous study, three focused on investigating replication challenges in various types of surveys, two provided detailed documentation and descriptions to aid future replication, and one recommended their study for future replication. Of the 10 replicated papers, 33% fo-

2771-1110/24/$31.00 ©2024 IEEE
DOI 10.1109/ISMAR-Adjunct64951.2024.00020

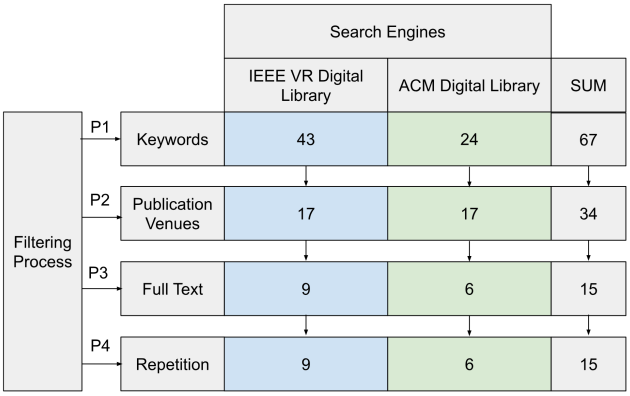

Figure 2: Literature review process

Table 1: Summery of surveyed papers

Refrence	Year	Published Venue	XR	Replication of
[1]	2023	ISMAR-Adjunct	VR	[29]
[24]	2023	CSCW	VR	[9]
[19]	2022	IEEE VRW	AR	[20]
[16]	2022	TVCG	VR	[17]
[26]	2022	ISMAR-Adjunct	AR	[11]
[10]	2022	TVCG	VR	[3]
[28]	2021	CHI	VR	[5, 30]
[21]	2021	CHI	VR	[18]
[2]	2020	IEEE VRW	AR	[19]

cused on AR, while 67% utilized VR in their replication methods. Table 1 provides a summary of these ten replicated papers and papers that were replicated from.

3 Replication Purposes

To address **RQ1**, we explored the underlying reasons for replicating previous studies. We have summarized these reasons as follows:

- Testing with different output displays: Arefin et al., replicated a previous study conducted with the Microvision Nomad display [12] using a custom-built AR haploscope[2]. They aimed to understand if the findings of the original study can be broadly generalized to optical see-through AR user interfaces.

- Verifying results avoid bias: Another reason for replicating a study is to verify whether the results of the original study are free from bias. As an example, in a previous study, the authors replicated a prior study by [20] to investigate if the virtual objects are biased in the direction of the dominant eye. Their findings confirmed that perceptual accuracy is not biased in that direction [19].

- Adopting solutions from other fields: Noah et al. adopted a security method, the Things scheme [9], in both AR/VR domains [24]. Similarly, Fennedy et al. adopted OctoPocus [3], to 3D mid-air gestures in VR [10].

- Investigating different features: Xu et al. recommended the replication of their study in different conditions such as using objects with a wider diversity of properties, including varying sizes and shapes to investigate spatial memory in VR [32]. By incorporating a broader range of object characteristics, researchers can better understand how different features impact spatial memory. Phillips et al. replicated the original study proposed by [11] in both outdoor and indoor settings. They modified some aspects of the original methodology, including the location, distance, and target [26].

- Enhancing understanding: when the findings of a study are not comprehensive or no conclusive results are achieved, replication of study can help understanding better: Andrich et al. replicated a prior study and added new variables to examine the effects of workload and sun speed on time perception. While they found that cognitive workload influenced time perception in VR, consistent with the original study, they did not observe significant differences related to virtual sun movements [1].

- Exploring a different target community: Noah et al. replicated a prior study [9] to perform a comparative analysis of cross-cultural differences between participants from Germany and the USA. They aimed to explore how cultural factors, particularly privacy perceptions, influence the acceptance and effectiveness of the authentication scheme [24].

- Exploring a different experiment environment: Kelly et al. (2022) replicated a lab-based experiment on locomotion interfaces in a remote setting [16]. The results collected remotely mirrored those found in their prior lab study (2020) [17].

- Testing generalizability: by replicating the study in various environments, researchers can assess the consistency and generalizability of the original results. Tran et al., [28] investigated the generalizability of a prior study [5] by replicating the study across different environments (remote).

4 Methods to Facilitate Replication

Hepperle et al. adapted solutions from other fields to address the issues of replicability in HCI and VR [15]. Previous studies have provided various levels of documentation and information to facilitate future replications [6, 22]. However, many of these studies fall short of offering comprehensive documentation and details necessary for precise replication. Here, we summarize some methods that can improve the accuracy of replications. Efforts to facilitate the replication of a study fall on two main groups: the authors and developers of the original study, and the publishers and venues where the research is disseminated. This section focuses on strategies to address **RQ2**.

4.1 Authors and Developers Roles

- **Employing open-source toolkits and assets:** Using a common toolkit standardizes the tools and methods used across different studies [15]. This standardization can improve the comparability of results from different replications. Using open-source standardized assets allows researchers to exactly replicate similar graphical components which can decrease biases. In addition, open-source toolkits often come with documentation and example code. This can help the procedures followed during replication are consistent with the original study.

- **Providing open-source codes:** Sharing open-source codes can facilitate replication of the results by supplying all the necessary functions and data access for others to follow the same steps [13]. Several common repositories that have been used for sharing open source codes are GitHub[1], GitLab [2] [15], Zenodo [3] [6], and OSF [4] [28].

[1]https://github.com/
[2]https://about.gitlab.com/
[3]https://zenodo.org/
[4]https://osf.io/c

- **Sharing test environments:** Even if the open-source version of the application is not available to future developers, providing a complete test environment can still help other researchers to replicate the conditions and methods of the original study. This can facilitate better understanding and testing of the environment. The running application should be actively executing and delivering functionality on a AR/VR device. In addition to online repositories (eg., GitHub, GitLab, or Zenodo), the application can be shared through platforms that support the distribution of VR content such as Unity Asset Store [5], Unreal Engine Marketplace [6], Meta store [7], itchio [8], and Viveport [9].

- **Transparent description of methodology and workflow:** Scientists should attempt to transparently describe their methodology. Miksik et al. provide full technical details of their system to aid replication [22]. In addition to providing a description, illustrating the workflow of the XR application can further facilitate replication [23].

- **Documenting the recorded Data:** Hepperle et al. highlighted the importance of preparing recorded data for further analysis using statistical software. They argue that proper data preparation is essential for ensuring that subsequent analyses are accurate and reliable [15]. The data to be recorded could include raw, encrypted, and anonymized experimental data, observational data from the experiment, databases, pre-questionnaires and post-questionnaires, recruitment data, data processing details, analytical methods, interview data, and ethical considerations, among other relevant information.

- **Employing standardized questionnaires:** Using standard questionnaires for the subjective evaluation of an experiment not only enhances the study's reliability but also facilitates easier replication. The finding from a survey by Tran et al. showed that questionnaires are the dominant measurements in XR research but the lack of standardized measurement has led to problematic trends. Specifically, the frequent use of non-validated or custom questionnaires limits the comparability of results, potentially contributing to a replication crisis in this growing field [31]. In addition, standardized questionnaires are widely recognized and familiar to many developers, which helps in maintaining consistency across different studies. Platforms such as HCI Studies [10], SoSci Survey can help in finding common questionnaires [11] [15].

- **Documenting questionnaires:** Simone et al. provided questionnaires and all forms used to collect feedback from users to allow replication in similar experiments [6].

- **Semi-structured interviews:** Conducting semi-structured interviews can aid future replication efforts [21]. By capturing nuanced information that might not be evident in written reports, semi-structured interviews help ensure that future researchers understand the original study's processes and decisions more thoroughly. This understanding can lead to more accurate and reliable replications, as researchers can account for variables and considerations that might otherwise be overlooked. several previous studies used semi-structured interviews to collect users' feedback and documented this data [6].

[5]https://assetstore.unity.com/
[6]https://www.unrealengine.com/marketplace/en-US/store
[7]https://www.meta.com/experiences/
[8]https://itch.io/
[9]https://www.viveport.com/filter-page/app?hl=en-US
[10]https://hci-studies.org/
[11]https://www.soscisurvey.de/

Figure 3 illustrates a step-by-step process designed to facilitate more accurate and precise replication.

4.2 Publishers Role

- **Dedicating tutorials and Workshops:** One reason authors might not make their work replicable is a lack of awareness about the importance of replication. Tutorial sessions or workshops dedicated to replication studies can help address this. These sessions can educate researchers on replication methodologies, challenges, and best practices, and provide hands-on opportunities to learn about replication tools and techniques. The limited number of studies that provided detailed documentation underscores the critical importance of comprehensive reporting.

- **Providing reliable repositories:** Providing reliable repositories in collaboration with publishers is crucial for maintaining the accessibility of open-source software and other research materials. Some papers include their software, open-source code, and additional resources, but over time, many repositories may become inaccessible or unsupported. To address this issue, using well-established repositories that have long-term support and stability can ensure that these resources remain available and useful.

- **Showcase Successful Replications:** Highlight successful replication studies in the conference program. This can include dedicated sessions or special awards for replication efforts, showcasing their importance and providing a model for others

5 Conclusion

This paper reviewed 15 studies related to replication in XR, highlighting the need for comprehensive documentation and standardized methodologies to facilitate precise replications. We identified various purposes for replication in HCI, such as testing with different displays, verifying results to avoid bias, adopting solutions from other fields, and investigating different environments and features. The findings indicate that employing open-source toolkits, providing detailed documentation, and ensuring accessible repositories are crucial for enhancing replicability. Future research should focus on creating a robust framework for replication in XR HCI to ensure the reliability and validity of scientific findings in this rapidly evolving field.

Acknowledgments

We like to thank and acknowledge the funding support provided by the National Science Foundation via grants NSF 2238313.

References

[1] A. Andrich, F. Weidner, and W. Broll. Zeitgebers, time judgments, and vr: A constructive replication study. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 1–2. IEEE, 2023. 1, 2

[2] M. S. Arefin, N. Phillips, A. Plopski, J. L. Gabbard, and J. E. Swan. Impact of ar display context switching and focal distance switching on human performance: Replication on an ar haploscope. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 571–572. IEEE, 2020. 2

[3] O. Bau and W. E. Mackay. Octopocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pp. 37–46, 2008. 2

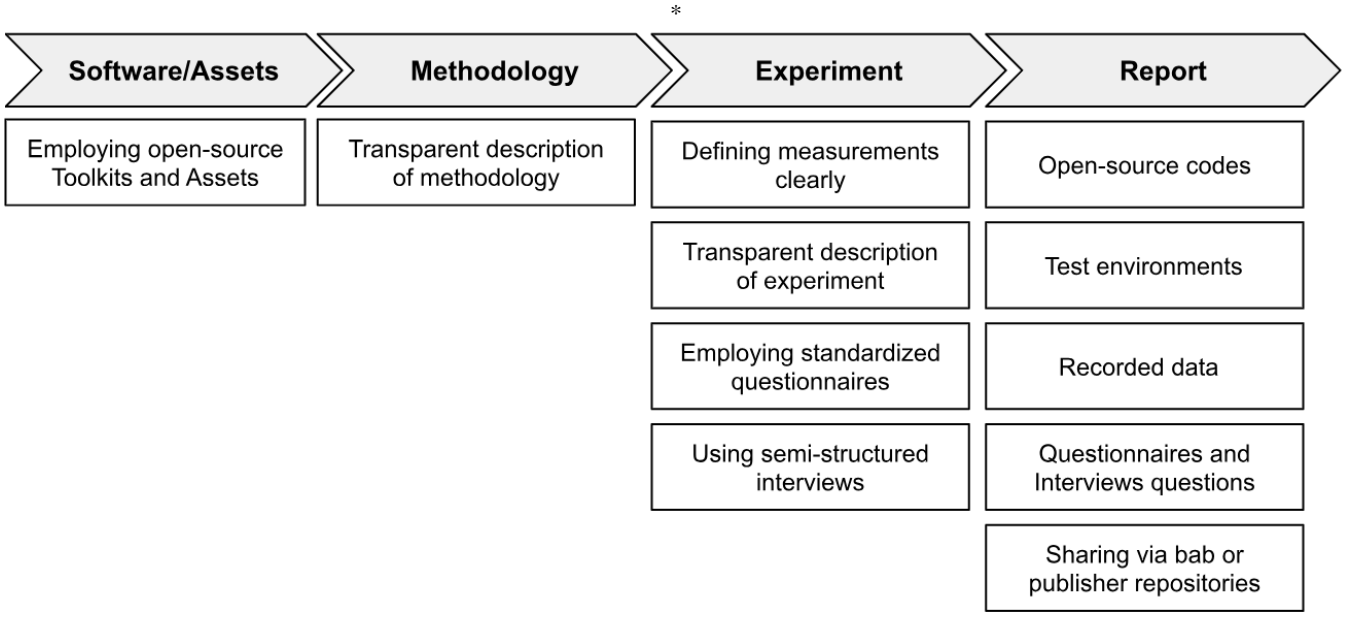

Figure 3: Steps to enhance the replication process

[4] Z. Borhani, P. Sharma, and F. R. Ortega. Survey of annotations in extended reality systems. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 1

[5] L. D. Clark, A. B. Bhagat, and S. L. Riggs. Extending fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies*, 139:102413, 2020. 2

[6] F. De Simone, J. Li, H. G. Debarba, A. El Ali, S. N. Gunkel, and P. Cesar. Watching videos together in social virtual reality: An experimental study on user's qoe. In *2019 IEEE Conference on virtual reality and 3d user interfaces (VR)*, pp. 890–891. IEEE, 2019. 2, 3

[7] M. Derksen and J. Morawski. Kinds of replication: Examining the meanings of "conceptual replication" and "direct replication". *Perspectives on Psychological Science*, 17(5):1490–1505, 2022. 1

[8] A. Dünser, R. Grasset, and M. Billinghurst. *A survey of evaluation techniques used in augmented reality studies*. Human Interface Technology Laboratory New Zealand, 2008. 1

[9] R. Düzgün, P. Mayer, and M. Volkamer. Shoulder-surfing resistant authentication for augmented reality. In *Nordic Human-Computer Interaction Conference*, pp. 1–13, 2022. 2

[10] K. Fennedy, J. Hartmann, Q. Roy, S. T. Perrault, and D. Vogel. Octopocus in vr: Using a dynamic guide for 3d mid-air gestures in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 27(12):4425–4438, 2021. 2

[11] S. S. Fukusima, J. M. Loomis, and J. A. Da Silva. Visual perception of egocentric distance as assessed by triangulation. *Journal of experimental psychology: Human perception and performance*, 23(1):86, 1997. 2

[12] J. L. Gabbard, D. G. Mehra, and J. E. Swan. Effects of ar display context switching and focal distance switching on human performance. *IEEE transactions on visualization and computer graphics*, 25(6):2228–2241, 2018. 2

[13] R. F. Gramaccioni, C. Marinoni, C. Chen, A. Uncini, and D. Comminiello. L3das23: Learning 3d audio sources for audio-visual extended reality. *IEEE Open Journal of Signal Processing*, 2024. 2

[14] W. He, B. Swift, H. Gardner, M. Xi, and M. Adcock. Reducing latency in a collaborative augmented reality service. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–9, 2019. 1

[15] D. Hepperle, T. Dienlin, and M. Wölfel. Reducing the human factor in virtual reality research to increase reproducibility and replicability. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 100–105. IEEE, 2021. 2, 3

[16] J. W. Kelly, M. Hoover, T. A. Doty, A. Renner, M. Zimmerman, K. Knuth, L. A. Cherep, and S. B. Gilbert. Remote research on locomotion interfaces for virtual reality: Replication of a lab-based study on teleporting interfaces. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2037–2046, 2022. 2

[17] J. W. Kelly, A. G. Ostrander, A. F. Lim, L. A. Cherep, and S. B. Gilbert. Teleporting through virtual environments: Effects of path scale and environment scale on spatial updating. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1841–1850, 2020. 2

[18] M. Khamis, L. Trotter, V. Mäkelä, E. v. Zezschwitz, J. Le, A. Bulling, and F. Alt. Cueauth: Comparing touch, mid-air gestures, and gaze for cue-based authentication on situated displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–22, 2018. 2

[19] F. A. Khan, M. S. Arefin, N. Phillips, and J. E. Swan. A replication study to measure the perceived three-dimensional location of virtual objects in optical see through augmented reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 796–797. IEEE, 2022. 2

[20] F. A. Khan, V. V. R. M. K. Rao, D. Wu, M. S. Arefin, N. Phillips, J. E. Swan, et al. Measuring the perceived three-dimensional location of virtual objects in optical see-through augmented reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 109–117. IEEE, 2021. 2

[21] F. Mathis, K. Vaniea, and M. Khamis. Replicueauth: Validating the use of a lab-based virtual reality setup for evaluating authentication systems. In *Proceedings of the 2021 chi conference on human factors in computing systems*, pp. 1–18, 2021. 2, 3

[22] O. Miksik, V. Vineet, M. Lidegaard, R. Prasaath, M. Nießner, S. Golodetz, S. L. Hicks, P. Pérez, S. Izadi, and P. H. Torr. The semantic paintbrush: Interactive 3d mapping and recognition in large outdoor spaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3317–3326, 2015. 2, 3

[23] C. Neamțu, R. Comes, D.-M. Popovici, E. Băutu, M.-S. Liliana, A. Syrotnik, and M.-I. Popovici. Evaluating user experience in the context of cultural heritage dissemination using extended reality: A case study of the dacian bronze matrix with hollow design. *ACM Journal on Computing and Cultural Heritage*, 17(2):1–21, 2024. 3

[24] N. Noah, P. Mayer, and S. Das. A proposal to study shoulder-surfing

resistant authentication for augmented and virtual reality: Replication study in the us. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, pp. 317–322, 2023. 2

[25] P. Patil, R. D. Peng, and J. T. Leek. A statistical definition for reproducibility and replicability. *BioRxiv*, p. 066803, 2016. 1

[26] N. Phillips, F. A. Khan, M. S. Arefin, C. L. Bethel, J. Stetanucci, and J. E. Swan. A conceptual replication and extension of triangulation by walking for measuring perceived distance through a wall. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 278–282. IEEE, 2022. 2

[27] D. E. Rzig, N. Iqbal, I. Attisano, X. Qin, and F. Hassan. Virtual reality (vr) automated testing in the wild: A case study on unity-based vr applications. In *Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis*, pp. 1269–1281, 2023. 1

[28] D. Saffo, S. Di Bartolomeo, C. Yildirim, and C. Dunne. Remote and collaborative virtual reality experiments via social vr platforms. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–15, 2021. 2

[29] C. Schatzschneider, G. Bruder, and F. Steinicke. Who turned the clock? effects of manipulated zeitgebers, cognitive load and immersion on time estimation. *IEEE transactions on visualization and computer graphics*, 22(4):1387–1395, 2016. 2

[30] A. Tang, M. Tory, B. Po, P. Neumann, and S. Carpendale. Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1181–1190, 2006. 2

[31] T. Q. Tran, T. Langlotz, and H. Regenbrecht. A survey on measuring presence in mixed reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–38, 2024. 1, 3

[32] M. Xu, M. Murcia-López, and A. Steed. Object location memory error in virtual and real environments. In *2017 IEEE Virtual Reality (VR)*, pp. 315–316. IEEE, 2017. 2