

A Systematic Review and Taxonomy of VR Usability Heuristics: Trends, Challenges, and Directions

MOHD KAMAL OTHMAN

Cognitive Science, Universiti Malaysia Sarawak Fakulti Sains Kognitif dan Pembangunan Manusia, Kota Samarahan, Sarawak, Malaysia, omkamal@unimas.my

RAHIMAH MAT

Cognitive Science, Universiti Malaysia Sarawak Fakulti Sains Kognitif dan Pembangunan Manusia, Kota Samarahan, Sarawak, Malaysia, 21010206@siswa.unimas.my

IPG, Institut Pendidikan Guru Kampus Batu Lintang Kuching, Kuching, Malaysia, 21010206@siswa.unimas.my

NOREHAN ZULKIPLY

Cognitive Science, Universiti Malaysia Sarawak Fakulti Sains Kognitif dan Pembangunan Manusia, Kota Samarahan, Sarawak, Malaysia, znorehan@unimas.my

As Virtual Reality (VR) becomes prevalent in high-stakes and everyday applications, ensuring seamless user experiences (UX) is essential. While Heuristic Evaluation (HE) is a highly valuable usability inspection method, its application across VR domains and devices remains fragmented and underexplored. This review examines VR usability research from 2004 to 2025, drawing on 53 rigorously selected studies. We propose a methodological taxonomy categorising heuristic approaches into four distinct types: VR-Specific and Custom Heuristics (41.5%), Combinations and Mixed-Methods (26.4%), Traditional/General Heuristics (24.5%), and Unspecified evaluations (7.6%). While Core HCI and Technology remain the foundational domain, HE has rapidly expanded into Education, Industrial Training, and Healthcare. Hardware usage predominantly involves immersive Wearables (HMDs), with hybrid 2D/3D systems emerging strategically in clinical and industrial contexts. Crucially, traditional 2D heuristics often fail to capture VR's unique spatial, physical, and cognitive complexities, including simulator sickness, multi-user latency, and spatial UI clutter. Recent literature (2024–2025) reveals a critical shift toward mixed-method evaluations, pairing HE with standardised metrics such as the System Usability Scale (SUS), and validated, domain-specific frameworks (e.g., for wearable VR educational apps). By addressing these gaps, adopting mixed methods, and refining frameworks, the community can deliver intuitive, safe, and user-centred VR experiences, unlocking its full transformative potential.

CCS CONCEPTS • Human-centered computing • Human-centered computing~Human computer interaction (HCI)~HCI design and evaluation methods~Heuristic evaluations

Additional Keywords and Phrases: Virtual Reality, VR usability taxonomy, User Experience, Taxonomy of Heuristic Evaluation

1 INTRODUCTION

Heuristic evaluation (HE) has become a cornerstone of usability assessment in human-computer interaction (HCI) and user experience (UX) design. It systematically applies predefined usability principles (heuristics) through expert evaluators to efficiently identify usability issues and optimise design [1, 2]. This translates to more user-friendly and impactful experiences across diverse interfaces and technologies [3, 4, 5].

Before discussing the adaptation of HE to Virtual Reality (VR), it is essential to establish its conceptual foundation in traditional HCI. HE, pioneered by Jakob Nielsen and Rolf Molich in 1990, is a widely used heuristic usability inspection method in which evaluators examine an interface against recognised usability principles (the "heuristics"). Traditional HE by Nielsen (1994) typically involves 3–5 expert evaluators inspecting interfaces against 10 usability principles (e.g., visibility of system status, user control and freedom), rating severity, and identifying issues early in the design process. These classic heuristics were designed to assess traditional 2D desktop and mobile interfaces. While highly effective for flat-screen graphical user interfaces (GUIs), traditional HE frameworks inherently lack the dimensions needed to evaluate the spatial, physical, and immersive characteristics of Extended Reality environments. Factors such as 3D navigation, presence, embodiment, simulator sickness, and cognitive load demand modified or entirely new heuristics to capture these unique experiential aspects.

HE's cost-effectiveness and ability to identify usability concerns early in development make it particularly valuable for emerging technologies such as VR [6]. In the unique context of VR, HE extends its reach beyond traditional interfaces, adapting to assess the usability and UX of immersive digital environments [7, 8]. Experts evaluate VR systems against usability principles that address VR's challenges and opportunities, such as navigation in three-dimensional spaces and interaction with virtual objects [9].

However, the richness of VR's functionalities can also become a double-edged sword. Usability issues can easily arise, compromising UX and immersion [10, 11] across increasingly diverse hardware, software, and applications in various fields [12, 13]. Usability and UX are intertwined in VR, influencing how users interact with and enjoy the virtual world [14, 15], while ensuring smooth navigation, effortless engagement, and a compelling and believable alternate reality [16, 17].

VR usability encompasses interface design, control schemes, and interaction paradigms, all aimed at making users feel comfortable and efficient as they accomplish their goals [9]. It also prioritises user satisfaction, efficiency, and effectiveness within the virtual environment (VE) [3]. When well-executed, VR usability seamlessly blends users into the digital world, enhancing their satisfaction and immersion [18, 19].

Conversely, UX focuses on perceptions and emotional responses within the VR environment [20]. It encompasses navigation mechanics, visual and audio realism, and the system's responsiveness to user movements [21]. A positive VR UX is characterised by users feeling immersed and interacting effortlessly while avoiding discomfort or motion sickness [22].

Proactively addressing usability issues before user exposure is crucial for maximising VR's potential [23]. Using established guidelines or heuristics enables developers to optimise the usability and UX of VR applications, creating a more enjoyable and effective VR experience for users.

This research investigates specific heuristics employed in previous VR usability studies. We examine the diverse domains in which HE has been applied and the devices used across settings. We

also identify potential challenges when applying general heuristics to VR evaluations and review solutions proposed by the authors in relevant studies. We then analyse research trends and suggest areas for further investigation concerning VR systems. This study is guided by five primary research questions, which are presented in Table 1. The methodology is outlined in the following section, followed by the presentation of Results and Discussion, in which we address our five primary research questions. Finally, we conclude our study and offer suggestions for future work and its potential implications.

Table 1: Research Questions for the Systematic Review

Research Questions	
RQ1	What specific heuristics have been employed in previous VR usability studies?
RQ2	Which domain has HE been applied to in the study of VR?
RQ3	What devices have been used in these VR usability studies employing HE across different domains?
RQ4	<i>RQ4a:</i> What challenges have been encountered when applying general-purpose heuristics to VR evaluations compared to VR-specific heuristics? <i>RQ4b:</i> What solutions have been proposed by authors in relevant studies to address the challenges identified in RQ4a?
RQ5	Based on the identified trends in existing VR HE studies, what research areas hold the most potential to contribute effectively to improving usability and UX in VR systems?

2 METHODS

We conducted a thorough literature review, following the guidelines established by [24] and [25]. The procedure was as follows.

2.1 Data Sources

We carefully selected four electronic databases for this review: SCOPUS, IEEE Xplore, ACM Digital Library, and Web of Science. These databases were chosen for their renowned quality, extensive coverage of relevant research across information systems, education, technology, engineering, computer science, and related fields, their regular publication of articles on VR and usability evaluation, and their ready accessibility to researchers. This combination ensured comprehensive, high-quality, and accessible data sources for our systematic literature review to address the RQs.

2.2 Search Strategy and Terms

To identify relevant studies precisely while minimising irrelevant results, we utilised a combination of primary search terms ("heuristic evaluation" and "virtual reality") and refined search strings (Table 2). For instance, string 3 limited the search to articles published between 2004 and February 2026, focusing on significant VR development and HE applications.

Table 2: Search strings.

No	Search strings
1	heuristic AND evaluation AND virtual AND reality
2	("All Metadata": heuristic evaluation) AND ("All Metadata": virtual reality)
3	[All: heuristic evaluation virtual reality] AND [E-Publication Date: (01/01/2004 to 26/02/2026)]

2.3 Study selection criteria

We defined strict inclusion and exclusion criteria (Table 3) to ensure a focused, high-quality selection.

Table 3: Inclusion and Exclusion Criteria.

Inclusion criteria	Exclusion criteria
Studies explicitly applying heuristic evaluation for VR system evaluation.	Theses, abstracts, book chapters, reports, SLR/Review articles/mapping studies, and lecture notes. Exclusion of grey literature was applied to prioritise rigorously peer-reviewed sources with established methodological standards, consistent with common practices in HCI/CS/VR systematic reviews and resource constraints for full-text appraisal of lengthy documents
Studies involving VR systems.	Duplicate papers.
Research articles published between 2004 and 2026 (cut off 26 th February 2026).	Other technology – MR/XR/AR. While Rauschnabel et al. [90] position Extended Reality (XR) as an inclusive umbrella encompassing VR, AR, and MR, they emphasise fundamental differences between the immersive nature of VR and the physical-world overlays of AR. To ensure conceptual precision in this systematic literature review, studies primarily framed as general XR or AR were excluded unless their primary focus and evaluation explicitly centred on immersive VR systems.
Papers written in English.	Other evaluation methods – not heuristic evaluation Papers not written in English.

2.4 Quality Assessment

We implemented a two-step quality assessment process to ensure the review included high-quality and impactful studies. First, two researchers independently screened the full-text articles of all studies meeting the initial inclusion criteria. Second, any discrepancies in selection were discussed and resolved collaboratively to guarantee consistency and objectivity. Ultimately, only articles published in peer-reviewed journals exhibiting rigorous methodology were included in the final review.

2.5 Data Extraction

For each included study, we meticulously extracted relevant information regarding eight key elements: (1) Author(s), (2) Heuristic(s) used in the VR system evaluation, (3) Domain of the study, (4) Devices used in the study, (5) Issues associated with the heuristic(s) and proposed solutions, (6) Trends identified in the VR HE studies, (7) Potential avenues for future research in VR HE, and (8) Limitations.

2.6 Data synthesis

The extracted data were then organised and analysed based on the specific domain of each study. This structured approach enabled us to identify patterns, trends, limitations and key insights across different VR contexts. Table 5 in the results section summarises the extracted information from each study, providing a readily accessible overview of the findings.

3 STUDY SELECTION

3.1 Identification and Screening

The initial search across the four selected databases yielded 635 potentially relevant articles. These articles were then subjected to a rigorous screening process using predefined inclusion and exclusion criteria. Figure 1 depicts the number of papers identified, screened, and included at each stage, following PRISMA guidelines [25].

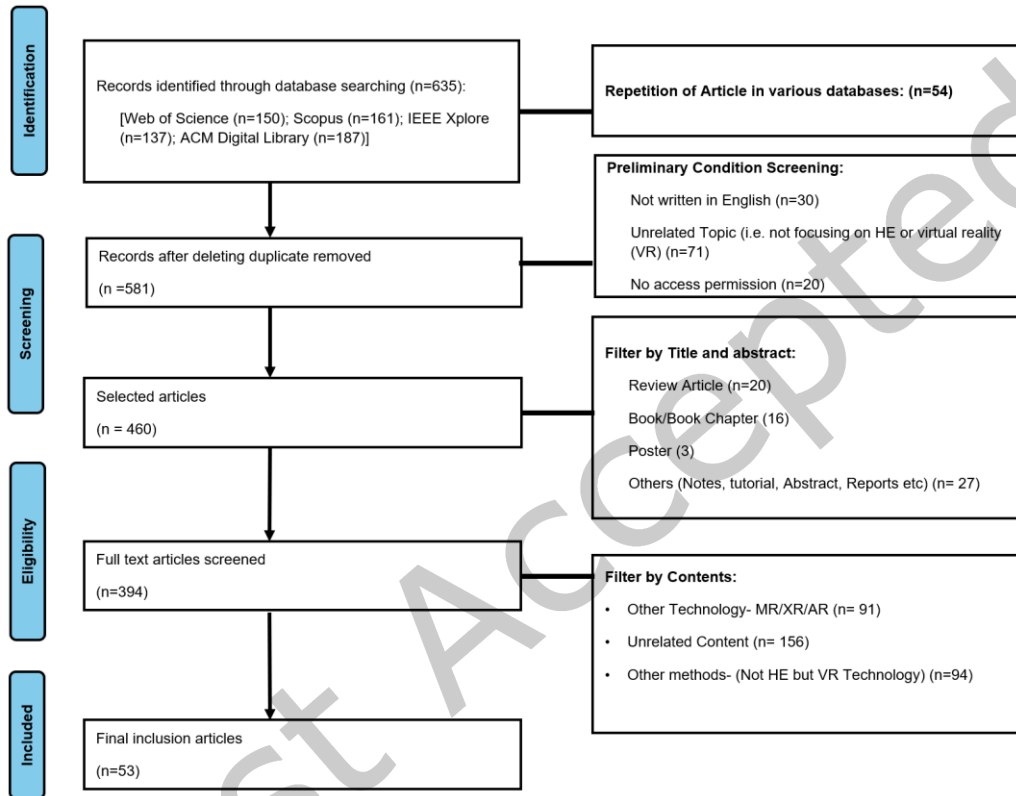


Figure 1: The Article Selection Process

3.2 Summary of Included Studies

Following a meticulous screening and exclusion process aligned with PRISMA guidelines [25], we ensured a rigorous, transparent review through distinct stages of identification, screening, and qualitative analysis (Figure 1). Two independent researchers conducted quality assessments and eligibility screening using pre-defined inclusion and exclusion criteria (Table 3). Any discrepancies were resolved through consensus discussions. The detailed screening process involved:

Initial Screening: A total of 635 records were identified across various databases. Duplicate entries were removed, resulting in 581 unique records. Following this, twenty (20) articles were excluded due

to access restrictions, thirty (30) due to language barriers (non-English), and seventy-one (71) for being irrelevant to the topic or focus area.

Title and Abstract Screening: The remaining 460 records were reviewed for alignment with the established inclusion and exclusion criteria. This step further excluded sixty-six (66) articles.

Content Screening: A total of 394 full-text articles were assessed for eligibility. Ninety-one (91) articles were excluded for focusing on non-VR technologies, 156 for unrelated content, and ninety-four (94) for not evaluating HE within their scope.

Table 4: Proposed Taxonomy of Heuristic Evaluation Approaches in VR Usability Studies (N=53)

Classification Category	Subcategory / Framework	Frequency	Percentage %
1. VR-Specific & Custom Heuristics (Highly tailored for immersive 3D spatial environments)	Sutcliffe & Gault (12 VR Heuristics) [7]	9	41.50%
	126 Guidelines [9]	2	
	16 Virtual World Design Heuristics [37]	2	
	WVREA Framework (Wearable VR Educational Apps) [89]	1	
	Guo et al. (9 VR Design Heuristics) [11]	1	
	Other Custom/Domain-Specific (e.g., Playability PHEG, Path Planning)	7	
2. Combinations & Mixed-Methods (Pairing heuristics with quantitative/qualitative metrics or merging multiple sets)	HE + Standardised Questionnaires (e.g., SUS, Presence Questionnaire, Simulator Sickness)	8	26.4%
	Combined Heuristic Sets (e.g., Nielsen + Sutcliffe, or Nielsen + Munoz)	6	
	Nielsen's 10 Heuristics [6, 30]	11	
3. Traditional / General Heuristics (Originally designed for 2D graphical user interfaces)	Olsen's Heuristic [32]	1	24.5%
	Twelve Usability Guidelines (Xerox + Nielsen)	1	
4. Not Specified / Vague	Evaluations that did not explicitly name the framework used.	4	7.60%
Total		53	100%

**Note on terminology: In this review, 'Combinations' refers to the synthesis of multiple heuristic checklists (e.g., merging general and VR-specific principles), whereas 'Mixed-Methods' refers to pairing Heuristic Evaluation with distinct supplementary evaluation methodologies, such as standardised questionnaires (e.g., SUS) or observational protocols (e.g., CTA).*

Fifty-three (53) studies met all the inclusion criteria and were deemed eligible for inclusion in the final systematic literature review. These studies consistently demonstrated the use of HE in assessing VR systems, aligning with the specific focus of this review. An in-depth analysis of these 53 articles was conducted to address the research questions outlined in Table 1. Each study was assigned a unique ID based on its specific domain, facilitating targeted analysis and synthesis of the extracted information

(Table 5). The next section presents the key findings and insights of this analysis, addressing each research question individually.

4 RESULTS

Table 4 summarises the classification and frequency of heuristics employed in VR usability studies. We propose this four-category methodological taxonomy as a foundational reference for classifying and comparing VR heuristic evaluation approaches, highlighting the shift from traditional/general methods toward VR-specific and mixed-method strategies. Furthermore, Table 5 summarises the selected records and analyses the findings according to several criteria, including evaluation methods, measures/heuristics, research domain, issues addressed, devices used, findings, and limitations/solutions. This aligns with the research questions addressed in this study.

Table 5: Summary Of Studies Included in the Systematic Review (N=53)

ID	Reference	Evaluation Method	Measures/Heuristics	Domain	Issues Addressed	Device Used	Findings	Limitations/Solutions
1	[26]	HE	Nielsen [6]	Technology (smart artefact)	Applicability of conventional usability methods to VR	HTC Vive Pro, Oculus Rift	Usability methods can be applied with adjustments for immersion fidelity. Further research is needed on representation fidelity.	A small sample size may not represent a broader VR user population.
2	[27]	HE	Two lightweight rate adaptation heuristics	Technology (VR video streaming)	Challenges in unlocking the full potential of adaptive tile-based VR streaming	HMD	Lightweight heuristics improve video quality. Challenges remain in conducting HE for VR streaming.	Open challenges in HE for VR streaming need further investigation.
3	[28]	HE	Nielsen [6]	HCI (Firefox Reality)	Usability issues in VR applications	Desktop	77% of usability feedback mapped to Nielsen's heuristics, indicating potential for VR applications. Community efforts are suggested to promote usability.	The study focused on a single VR application, limiting generalizability.
4	[29]	HE	Twelve usability guidelines from Xerox and Nielsen [6]	HCI (online virtual world)	The student's ability to identify and explain heuristic problems	Desktop	Students identified problems but struggled to explain them or differentiate them from false positives. Improved training is needed.	Improving training in applying heuristics and identifying false positives in VR evaluations is needed.
5	[31]	HE +Expert interview	Olsen's Heuristic [32]	HCI (automotive user interface analysis)	Transitions from desktop to VR, novice user experience	Desktop, VR tablet	Desktop-to-VR transitions and indirect context portals confused novices. Tutorials and hints are recommended for novices.	The study focused on automotive interfaces, limiting generalizability to broader VR

								applications.
6	[33]	HE + User-centred evaluation + comparative evaluation	126 guidelines [9]	HCI (texture discrimination in VE)	Accuracy of different methods for texture discrimination	Desktop, Phantom force feedback device, vibrotactile data glove, paper prototypes	The Vibrotactile Data glove is effective for detecting grating textures and identifying shapes	Limited generalizability due to focus on specific textures and tasks.
7	[34]	HE	Path Planning Heuristics [35]	HCI (crowd behaviour and path planning in VE)	The complexity of path planning for large crowds and emergencies	Desktop	Synchronous path planning, integration of collective behaviour, and enhancements of the A* algorithm are proposed.	Limited validation beyond simulations and conceptual models.
8	[36]	HE + Driver method + interview	16 Virtual World Design and Aesthetics Heuristics [37]	HCI (design and usability of VE on the 3DGEN platform)	Effectiveness of HE for the 3DGEN platform and identification of user-system discrepancies	Desktop	HE is effective with platform-specific heuristics. The driver method reveals discrepancies between conceptual and mental models.	A small sample size limits generalisability. Real-world validation is needed.
9	[38]	HE + Usage observation + questionnaire + interviews + CEM (semiotic engineering)	126 guidelines [9]	HCI (interaction design and usability of the HybridDesk system)	Interaction and usability issues in hybrid VE systems	Stereoscopy, head tracking, wand, WIMP interface, VR-Nav, VR-Manipulation	Execution problems (wand button manipulation), perception problems (system feedback). A tightly coordinated system design is crucial for hybrid VEs.	Focusing on a single system limits generalisability. Further validation with other hybrid VEs is needed.
10	[39]	HE	Nielsen & Molich [30]	HCI (evaluation of dynamic guidance systems at transit hubs)	Risks and considerations for using immersive VR in HE	HTC Vive Pro HMD, HTC Controller, HTC Lighthouse 2.0 trackers, high-performance notebook	Immersive VR is effective for HE with considerations for cybersickness, performance, and participant expertise.	Limited generalisability due to the specific environments and applications. More studies are needed on different domains and concepts.
11	[40]	HE + Questionnaire	Nielsen [6], Simulator Sickness Questionnaire [41], Presence Questionnaire [42]	HCI	Interactive problems, discomfort from virtual crowds, and cybersickness in visualising crowd flow in Nørreport Metro Station	VR headset	The prototype improved after addressing interactive and comfort issues. Teleportation reduced cybersickness without compromising presence.	Limited generalisability beyond the specific environment and task.
12	[43]	HE	Heuristic analysis of finger motions, release-heuristic function	HCI (virtual grasp release in whole-hand virtual grasping)	Incorrect triggers, unintentional finger motion, subjective artefact trade-offs	Mirror-based "fish tank" VR display	The spring model improved release speed and accuracy, which users prefer. Heuristic detection is	Limited generalisability due to specific applications

							important for grasp release.	and hardware.
13	[44]	HE	Hand manipulation heuristic for immersive cube system	HCI (direct manipulation using finger tracking in an immersive cube)	Not applicable	Immersive CUBE	Finger-tracked manipulation in immersive CUBE is feasible and closer to natural than a wand, but slower and less precise. Further research is recommended.	Limited generalisability due to specific environment and interaction methods.
14	[45]	HE +Formative evaluation	Twelve Heuristics [7]	Marine Archaeology (4th-century BC Mazotos shipwreck)	UX and understanding of an underwater archaeological site	Computer, 3D glasses, Xbox controller, VR CAVE	Improved understanding of shipwrecks and engagement with archaeological data.	Limited generalisability due to specific fields and applications.
15	[8]	HE	A new set of heuristics	VR	Lack of formal methods for evaluating VR usability. Challenges in adapting traditional usability evaluation methods to VR, given its unique characteristics. Reliance on heuristics from desktop and mobile interfaces that may not be fully suitable for VR	VR headset	Nine usability problems in VR systems have been identified through surveys of VR users. A new set of heuristics has been developed to address these VR-specific usability issues.	Limited generalisability due to the scope and specific systems used. Lack of quantitative data.
16	[46]	HE	Nielsen [6]	HCI (iVR application)	Usability issues in knowledge representation, navigation, field-of-view, and accessibility	HMD (Oculus Quest)	Identified usability issues, informing application development.	A small sample size limits generalisability. Further testing with diverse users is needed.
17	[47]	HE	Not specified	HCI (VR simulation of human senses from a fetal point of view)	User experience and engagement for pregnant couples	HMD with additional sensors	Proposed features and functionalities to improve user experience.	Limited user testing and generalisability. Ethical considerations for pregnant audiences.
18	[48]	HE	Twenty-five principles [49], Nielsen [6], Nielsen [50]	HCI (telepsychology web and VR system for older adults)	Usability for elderly users unfamiliar with computers and VR technology	Desktop, VR HMD	Identified usability issues for elderly users. Recommendations for design and development were provided.	Limited generalisability due to specific user populations and applications.
19	[51]	HE	Bias hypothesis of the heuristic-systematic model (HSM)	Psychology	Persuasive impact of spatial presence on cognitive processing	Oculus Quest, Desktop	High technological immersiveness and spatial presence increased video credibility and positive thoughts. Technological	Small sample size and composition

							immersiveness had a small-to-medium effect on the valence of thoughts.	
20	[52]	HE + Questionnaire	UX/UCD/VR principles, Flow State Scale (FSS), Inspection Method, and System Usability Scale (SUS)	Psychology (therapy-support VE for physical rehabilitation)	Lack of motivation in patients. Limited guidance for developers	Not mentioned	Focused on improving guidelines for designing therapy-support VR environments through expert feedback and user testing.	Limited scope. Guidelines are based on the single facility (DIF). Only focusing on physical and neglecting cognitive and neurological
21	[53]	HE	M-GRASP (Memory Greedy Randomised Adaptive Search Procedure)	Game (latency-aware partitioning in Distributed Virtual Environment (DVE) systems)	Issues with DVE and latency	Desktop	M-GRASP is an efficient heuristic method for solving latency-aware partitioning problems in DVE systems.	Limited evaluation scope. Also, it does not guarantee an optimal solution.
22	[54]	HE	A specific heuristic for a VR game	Game	Creating game design heuristics specific to VR	Oculus Rift Development Kit 1, PS3 controller	Proposed a set of VR-specific game design heuristics requiring academic evaluation.	The developed heuristics lack empirical validation.
23	[55]	HE	Twelve Heuristics [7], 16 Virtual World Design and Aesthetics Heuristics [37]	Game (evaluation of immersive low-cost VR in games)	Lack of instructions, hardware problems with Google Cardboard	Google Cardboard, Android OS	Recommended considering the application context and objective in HE. Suggested application characteristics, such as movement freedom or configuration options, are not universal rules. Provided recommendations and improvements based on reported errors.	Evaluated individual heuristic sets rather than comparing them. Context dependency, as the specific heuristics, influenced the performance of the heuristic sets.
24	[7]	HE	Twelve Heuristics [7]	Game	Overlap between the first three heuristics	HMD, hand-held interactive 3D mouse	Heuristics 8 and 9 (entry/exit points, consistent departures) were more applicable to desktop VR.	Acknowledge to tailor heuristics to different styles of VEs. Lacks comparison of evaluation methods.
25	[37]	HE	16 Virtual World Design and Aesthetics Heuristics [37]	Game	Lack of validated design heuristics for presence in VR. Inadequacy of existing VR usability evaluation methods	Desktop	16 Virtual World usability heuristics [35] worked better than Nielsen [6]. A balance between specificity and generality in heuristics is important.	Limited validation and more extensive user testing and evaluation are needed.
26	[56]	HE	Usability heuristics	Game	Challenges in Object-	Oculus Rift	The application could	Limited

			of VR applications		oriented programming teaching. Using VR to support learning.		support the teaching of the Object-Oriented Paradigm through Immersive Learning, but usability issues must be addressed.	scope, only focused on object-oriented VR games. Focus only on the basic concepts of object-oriented programming.
27	[57]	HE	Playability HE for Educational Computer Games – PHEG	Educational game	Communication skills deficiency in nursing practice and Limited opportunities for practising communication skills:	Laptop, Oculus Rift	The potential of using VR for nursing education.	Limited scenario complexity.
28	[18]	User testing + HE	HE: Nielsen et al. [30]; User Testing: Cognitive performance, Interaction performance	Game	Usability issues, discomfort with HMD, difficulties with keyboard navigation, and motion sickness	Desktop, Oculus Rift	Identified usability issues and proposed design improvements, including a "feet" icon for movement, different viewing options, emergency exit/pause, virtual clock, animated objects, multiuser option, and voice-controlled AI.	Limited context and comparison. It combines user testing and HE but relies heavily on self-reported data.
29	[58]	HE + SUS + Cognitive Walkthroughs	Nielsen & Molich [30], SUS	Education (virtual clinical simulation in health science education)	VR for medical training.	Oculus Rift	VE interesting, potential with changes. Recommendations for future HS education VE development. Future research on student use.	Only six experts were involved, limiting the generalisability of the findings.
30	[59]	HE + design guidelines	Various Heuristics	Education (navigation & wayfinding in 3D virtual learning spaces)	Navigation & wayfinding	Desktop	Derived heuristics and guidelines for designing 3D VR learning spaces for navigation and wayfinding	Only focusing on qualitative data
31	[60]	HE	Eight heuristics for student-centred VR lesson plans	Education (higher education - VR exploration of geomorphology theory)	Physical Geography students experience low-cost VR as an interactive learning platform in higher education.	HTC Vive	Developed heuristics for assessing VR. Insights into students' attitudes and behaviour towards VR in the classroom.	Focused on specific VR use cases and student populations. Did not measure learning outcomes.
32	[61]	HE + Presence Questionnaire+ Interview	Twelve Heuristics [7], Université du Québec en Outaouais Revised Presence Questionnaire (PQ)	Education (healthcare education - ALS-SimVR)	VR-based ALS training	Oculus Go Headset	VR-based ALS training is suitable, and a user-centred design approach is crucial for developing VR training.	Only five design considerations were highlighted. Also, it only

								involved small sample sizes.
33	[62]	HE + TAM questionnaire	Nielsen [6], Twelve Heuristics [7], TAM Questionnaire [63]	Education (Hajj learning procedures courseware)	VE as an alternative for training and learning	Desktop, mouse, keyboard	The applications required usability improvements. Older users are reluctant to use VR applications.	Not mentioned
34	[64]	HE	Twelve Heuristics [7]	Training (VR application for virtual manual assembly)	Navigation	Wii Remote, BalanceBoard	List of unsatisfactory aspects of the devices and their improvements	Limited to the devices used and future work needed to have additional devices
35	[65]	HE	Design-based Research (DBR) Heuristic	Training	Inadequacies in the Evaluation Framework for the VR Training Program HE	Desktop	The DBR model is suitable for designing and evaluating VR training artefacts for mine safety training. Model transferable and customisable. Criteria serve as design principles.	Technical barriers, such as the high cost of immersive VR and the computer skills of users
36	[66]	HE	Twelve Heuristics [7]	Training (crisis management)	Some problems did not fit well with the SG2004 heuristics	Head-mounted Device	Researchers adopt heuristics to suit their needs. A patchwork of heuristics is more useful than a single list. Further research on the patchwork approach is warranted.	Not mentioned
37	[11]	HE	Nielsen [6], Twelve Heuristics [7], Munoz et al. [67]	Training (VR training applications)	Challenges in identifying issues due to: - Heavy cognitive load of VR - Difficulty inspecting 3D environments - Body memories from VR experiences - Recording and describing issues - Locating understanding of reported issues - Establishing connections between issues	Oculus Quest 2 headset	Recommendations: Tailor heuristics for the VR context. Develop VR-specific HE support tools. Provide clear guidelines for recording and describing issues. Encourage mixed recording strategies. Conduct debriefing sessions to foster collaboration and understanding on issues.	Only focuses on single-type VR training applications.
38	[68]	HE	Twelve Heuristics [7]	Training	Safety training using a VR prototype	HMD	Need to consider insights from experienced VR.	Limited respondents
39	[69]	HE	Nielsen & Molich [30]	Tourism (Cultural heritage)	To provide a novel interaction for immersive exploration of an architectural site	Workbench (custom laser pointer, IR camera)	Novices are comfortable with interactions. Interface and setup are also intuitive and cost-effective. Further research on gyro-meter	The experiment was not conducted with the targeted users.

							effectiveness and spot detection accuracy.	
40	[70]	HE Questionnaire (HEQ)	User control/freedom, error handling/guidance	Tourism (Virtual Museum)	Evaluator expertise/time, end-user involvement	Smartphone	Improvements to user control/freedom and error handling/guidance are recommended.	Evaluator expertise/time, lack of end-user involvement.
41	[71]	HE + quantitative/qualitative measures	Twelve Heuristics [7]	Tourism (Virtual Museum tour for schools)	Navigation difficulties with a bad internet connection.	Head-mounted device (HMD)	Lack of skip/exit buttons, 2D/3D artefact distinction, edit menu/quiz feedback, and an informative museum map. Redesign focus on learning potential and assessment types.	Not mentioned
42	[72]	HE	Rusu et al. Twelve Heuristics [7] and MILE+	Tourism	Navigation, content, usability, user experience, and user testing issues	Google Cardboard	Navigation, content, usability, user experience, and user-testing issues were identified. New guidelines are proposed to improve the experience.	Limited sample, empirical research methods.
43	[73]	mHEAR checklist + SUS + HEQ +Cognitive Walkthrough	mHEAR checklist, SUS, HEQ,	Treatment (eating disorders - anorexia nervosa)	Infrastructural, technological, interoperability, financial, adoption problems; intervention delivery disagreement.	HTC VIVE VR system	Clear instructions are needed to avoid user confusion.	Not focusing on real users. Future work should test the VR application with real users.
44	[80]	HE	Sutcliffe & Gault (12 VR heuristics) [7]	K-12 Programming Education	Interface usability, excessive reading load, and intuitiveness of external controls in a VR learning environment.	Smartphones, VR BOX, wireless joysticks	No severe issues found; mostly inconvenient visual UI problems. External joysticks were unintuitive without prior explanation.	Limitation: Small sample size; subjective interpretations of 'presence'.
45	[81]	HE	Guo et al., 2024 (9 VR Design Heuristics) [11]	Collaborative Higher Education	VR motion sickness, steep learning curves for movement, and network latency across multiple users disrupt collaboration.	HTC VIVE, Oculus Quest	The system is highly effective for shared learning, but "Navigation" and "Visibility" scored lowest due to movement sickness and UI lag.	Future: Need custom heuristics for multi-user VR.
46	[82]	HE + Data Mining (Association Rules)	Dialogic teacher talk & Emotion coding schema	Teacher Education	Preservice teachers experiencing negative emotions (frustration/anxiety) due to uncertainty in achieving teaching goals.	OpenSimulator VR	AI-powered virtual students fostered positive emotions, inducing "reasonable frustration" that successfully challenged and improved pedagogical practice.	The study was designed as an exploratory case study involving only 22 preservice teachers, which limits

								the generalisability of the findings.
47	[83]	Modified HE (+/- Weighting)	Custom 14-Heuristic Set (Nielsen + Sutcliffe)	Nuclear Robotics	Cognitive overload from managing heterogeneous robot fleets, cluttered UI tags, and 3D navigation difficulties.	HTC Vive Pro, Flat-screen PCs, ROS	Identified three core needs: interface flexibility, individual robot tracking, and expanding sensor visualisation without clutter.	Limitation: Expertise bias among evaluators.
48	[84]	HE + Semi-structured Interviews	Nielsen (10 heuristics)	Additive Manufacturing	Visualising complex 3D internal defects, collaborative latency, and "depth errors" when drawing 3D annotations in mid-air.	HTC VIVE Pro, Meta Quest 3, VIVE XR Elite	Highly effective for collaborative inspection, but industrial experts (novices to VR) struggled initially with complex controller mappings.	Limitation: High hardware requirements.
49	[85]	Expert HE	Not mentioned	Cultural Tourism & Heritage	Visual clashes when merging 360 ° video with 3D-reconstructed assets; lighting mismatches; the "eye-level" illusion causing height disorientation.	Oculus Quest 2, 360° Cameras, LiDAR, MoCap	The gamified VR tour successfully integrated digital guides and reconstructions, though technical visual clashes required extensive mitigation.	Lack of End-User Testing. It was strictly an expert-based heuristic pre-screening. It was tested by a panel of only five domain experts.
50	[86]	HE + concurrent think-aloud (CTA)	14-item Game Usability Heuristics	Cultural Heritage Interpretation	Lack of emotional engagement; poor multimodal feedback; accurately representing spatial heritage without feeling "empty".	Meta Quest Pro	The interface and spatial layout scored high, but the accurate archaeological simulation felt emotionally empty without deep narrative integration.	Future: Implement AI-driven dynamic narratives and location-based audio.
51	[87]	HE + SUS + Nielsen Severity	Sutcliffe & Gault (12 heuristics) [7]	Healthcare / Telemedicine (MDTs)	Network latency lagging AI avatars; eye strain from reading complex clinical text on VR UI panels.	Meta Quest 3	High usability (SUS 72.7); AI avatars replacing text panels were highly effective, though latency degraded the sense of presence.	Limitation: Heavy network dependency; AI lacked deep clinical knowledge.
52	[88]	HE + Focus Groups	Subset of Nielsen's (Status, Control, Error Prevention)	Mental Health (Exposure Therapy)	Therapists are losing physical patient cues (sweat) in VR; patients are getting "performance anxiety" if the therapist uses a 2D monitor instead of VR.	Oculus HMDs, Desktop UI, 360° Video	A dual-interface is necessary: an HMD for the therapist builds trust with the child, but a Desktop UI is much easier to control the simulation.	Future: Integrating biometric wristbands to feed physiological patient data into the VR environment.
53	[89]	HE (multi-stage validation)	WVREA heuristics (comprehensive set for wearable VR educational apps) [94]	Education (Wearable VR in educational contexts)	Usability evaluation gaps in WVREA; need for pedagogical/immersion-specific heuristics	Wearable VR devices (e.g., HMDs)	Developed/validated a tailored heuristic framework addressing immersion, cognitive load, and pedagogy	Foundation for standardised WVREA usability;

								calls for further empirical testing in HE.
--	--	--	--	--	--	--	--	--

5 DISCUSSION

RQ1 What specific heuristics have been employed in previous VR usability studies?

Our review of the heuristics employed in previous VR studies reveals a diverse and highly fragmented landscape across the 53 documented evaluations. Rather than relying on a single standard, researchers utilise a wide array of frameworks. Table 4 and Figure 2 visually depict the classification and frequency of these methodologies. The data reveal four distinct approaches: the use of standalone VR-specific heuristics (41.5%, n=22), the integration of mixed-methods and combined heuristics (26.4%, n=14), the reliance on traditional 2D heuristics (24.5%, n=13), and studies in which the framework was not specified (7.6%, n=4). Sutcliffe and Gault's 12 VR Heuristics [7] emerged as the most widely used standalone framework, appearing in nine instances. This prominence suggests its strong fit with evaluating VR environments, as recognised by many researchers. However, the data also shows that many custom heuristics are scattered and rarely reused. Going beyond mere paper counts, this fragmentation occurs because general heuristics often fail to capture domain-specific spatial, cognitive, and physical complexities (such as simulator sickness, multi-user latency, or spatial UI clutter). Because existing heuristics are often too general or too rigidly tailored to a different niche, researchers are frequently forced to synthesise their own custom lists for each new study.

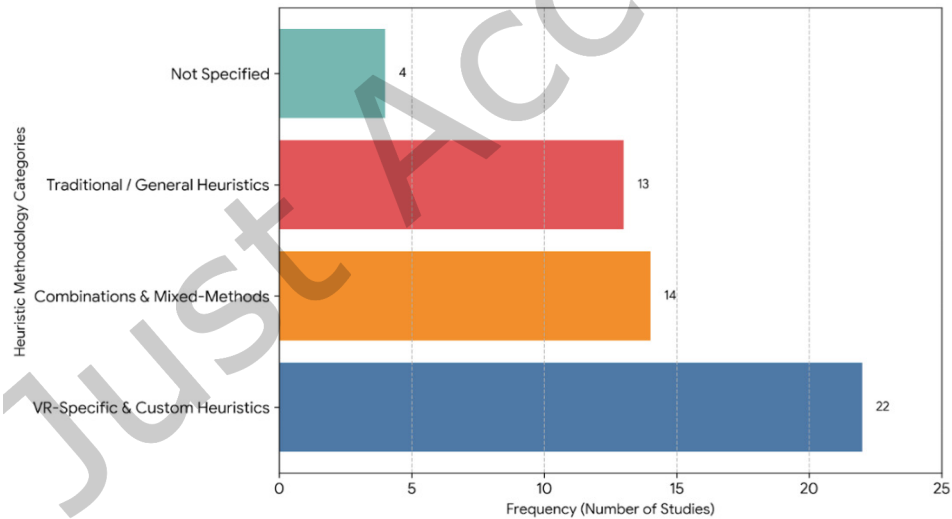


Figure 2: Distribution of heuristic evaluation methodologies across the reviewed VR studies (N=53)

The persistent fragmentation and low reuse of heuristics, with most frameworks appearing in only one or two studies, can be explained by three interlocking mechanisms, as is evident across the corpus. First, domain-specific affordances create a fundamental mismatch: general heuristics (e.g., Nielsen, Xerox) excel at assessing 2D visibility and error prevention but fail to capture VR's unique spatial navigation, embodiment, and presence dimensions. This failure drives researchers across domains, such as education [58, 60, 89], training [11, 37], and cultural heritage [85, 86] to abandon standard lists in favour of custom adaptations.

Second, rapid hardware evolution outpaces standardisation: early studies relied on desktop-to-HMD transitions (2004–2015), while studies from 2020–2025 reflect the dominance of standalone wearables (e.g., Meta Quest series) and hybrid setups. This hardware shift requires creating new custom heuristic lists for each device generation, as evidenced by the device-specific trends in Figure 3 and Table 6.

Third, a lack of validation infrastructure discourages reuse: unlike desktop HCI, the VR field currently lacks shared benchmark datasets or standardised severity scales, leading researchers to default to bespoke, single-use adaptations. These mechanisms explain why even promising, validated frameworks such as Sutcliffe and Gault [7] (used in 9 studies) remain outliers rather than becoming the universal standard. While Sutcliffe and Gault's heuristics emerged as the most frequently used [7], highlighting their suitability for VR evaluation, other influential heuristics, such as Nielsen's [6], were combined with both [7] and the twenty-five principles [49] in two studies, demonstrating their continued relevance through integration with others. However, Nielsen's framework [6] appears less frequently on its own, primarily appearing in combination with traditional heuristics such as the 12 Xerox usability checklist.

Further examination into the integration of multiple heuristics reveals interesting research trends. Four studies demonstrated the adaptability of Sutcliffe and Gault's heuristics [7] to diverse VR experiences by combining them with other frameworks. They were combined with Nielsen's [6] in two instances, highlighting a complementary relationship between general usability principles and VR-specific considerations. More importantly, while earlier studies occasionally paired HE with the System Usability Scale (SUS) [52, 58, 73], an analysis of recent literature (2024–2025) highlights a distinct shift toward mixed-method evaluations. To overcome the limitations of isolated heuristic checklists, researchers are now aggressively pairing HE with standardised quantitative and qualitative metrics. For example, recent evaluations heavily utilise the SUS [87], Data Mining [82], and Concurrent Think-Aloud (CTA) protocols [86] to accurately measure psychological factors such as immersion, emotional resonance, and physical fatigue.

However, the discussion around challenges encountered with general-purpose heuristics in VR (RQ4a) reveals their limitations in addressing domain-specific complexities and unique user experiences. For instance, early studies in education VR (e.g., [58]) highlight the inadequacy of existing frameworks in capturing pedagogical considerations and the immersive nature of VR learning environments. Similarly, research in VR games (e.g., [18]) emphasises the need for heuristics tailored to mitigate challenges such as multiple HMDs and natural language interactions, which traditional frameworks do not adequately address. In high-stakes environments, such as healthcare and industrial training, traditional heuristics also fail to account for the high cognitive load and the difficulty of

recording issues while interacting with immersive 3D environments [11], or the "operator blindspot" (the loss of physical patient cues when a therapist wears an HMD) [88].

The variety of identified heuristics, including those specifically designed for VR, underscores the dynamism of this research field. Fortunately, recent literature is actively addressing these historical gaps; for example, the introduction of the WVREA framework [89] provides a comprehensive, multi-stage validated heuristic set explicitly tailored for wearable VR educational apps, successfully merging pedagogical needs with immersive spatial metrics. Researchers are actively developing and adapting existing methodologies to address the unique challenges posed by VR environments. This continuous evolution bodes well for the future of VR usability, ensuring consistent improvement in user experience and effective interaction within VEs.

A recurring tension in the reviewed literature is the trade-off between heuristic specificity and generalisability. Highly tailored sets, such as the WVREA framework [89], Guo et al.'s 9 VR Design Heuristics [81], and custom domain-specific sets [83], achieve high diagnostic power in their target contexts (wearable education, collaborative learning, and nuclear robotics) but often sacrifice reusability across domains. Conversely, general heuristics (e.g., Nielsen [6], often combined with SUS [87]) are reused more frequently but consistently miss VR-unique phenomena, such as spatial clutter, latency-induced nausea, and the 'operator blind spot' [88].

Recent studies (2024–2025) have begun resolving this tension through hybrid strategies: pairing domain-specific heuristics with standardised metrics (SUS, CTA) or integrating AI extensions, such as generative avatars [87] and AI-driven virtual students [82]. While this mixed approach appears to balance fidelity and transferability, it simultaneously increases evaluation complexity and requires greater evaluator expertise. Future work should prioritise quantifying this trade-off empirically, for example, by comparing detection rates and false-positive ratios across different specificity levels in controlled HE scenarios, to guide practitioners toward scalable yet effective protocols.

RQ2 Which domain has HE been applied to in the study of VR?

Several domains in which HE has been applied in VR studies are illustrated in Figure 3. To avoid conceptual overlap, as all VR research inherently involves Human-Computer Interaction (HCI), we categorised studies by their specific application context. Studies that were domain-agnostic, focusing purely on fundamental UI mechanics, locomotion, or hardware evaluation, were grouped under Core HCI and Technology.

Core HCI and Technology emerges as the most researched area, with 17 studies solidifying its position as the foundation for VR usability exploration [8, 26, 27, 28, 29, 31, 33, 34, 36, 38, 39, 40, 43, 44, 46, 47, 48]. This robust interest underscores the inherent connection between fundamental interface principles and the unique complexities of spatial VR environments.

Eight studies investigate Games, focusing on the immersive world of VR gaming [7, 18, 37, 53, 54, 55, 56, 57]. This dedication reflects the ever-evolving landscape of VR entertainment and the constant pursuit of seamless user experiences within virtual worlds.

Nine studies examined the transformative potential of VR in Education [58, 59, 60, 61, 62, 80, 81, 82, 89]. This concerted effort highlights the growing recognition of VR's ability to revolutionise pedagogy and create engaging, interactive learning environments.

Tourism and Cultural Heritage research includes six studies highlighting the growing interest in virtual preservation and exploration [69, 70, 71, 72, 85, 86]. These investigations aim to optimise VR's use to showcase destinations, foster cultural understanding, and provide immersive historical journeys for travellers.

Healthcare, Psychology, and Treatment contribute to five studies [51, 52, 73, 87, 88]. These studies explore the impact of VR on behaviour and emotions, alongside recent expansions into high-stakes clinical applications, such as Virtual Reality Multidisciplinary Team (VRMDT) telemedicine [87] and mental health exposure therapy [88].

Training and Industrial Applications encompass seven studies [11, 64, 65, 66, 68, 83, 84]. These studies assess VR's effectiveness in skill development, simulation-based learning, and the recent surge in the operation of complex industrial digital twins, such as cyber-physical nuclear robotics [83] and additive manufacturing inspection [84].

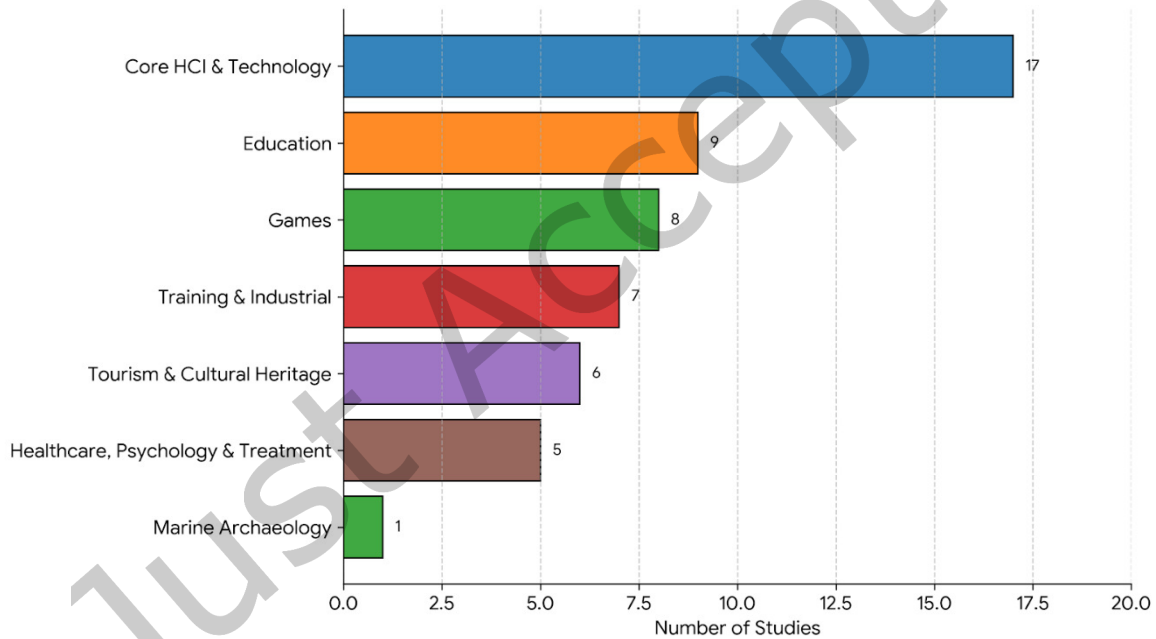


Figure 3: Distribution of VR usability studies across application domains (N=53)

Our analysis of the research domains (RQ2) in Figure 3 reveals a multifaceted VR research landscape, with diverse domains actively employing HE to optimise UX and unlock the full potential of this revolutionary technology. While Core HCI and Games established foundational heuristics, the recent expansion into high-stakes domains such as telemedicine and nuclear robotics underscores the

critical, expanding reach of VR. This comprehensive overview lays the groundwork for further investigation into the specific challenges and opportunities within each domain, aiming to refine VR usability methodologies and contribute to the continuous evolution of this dynamic field.

This diversity, however, also reveals historical gaps in domain-specific VR heuristics. As highlighted in the conclusion, VR education has historically relied on general heuristics and required more dedicated frameworks to account for the unique interaction modalities and pedagogical aspects of VR learning environments. Fortunately, recent literature has begun actively addressing these gaps. For example, the 2025 introduction of the WVREA framework [89] successfully provides a dedicated heuristic foundation for wearable VR educational apps. Addressing these remaining gaps through continued domain-specific heuristic development is crucial for ensuring effective and comprehensive evaluation of VR applications across various fields.

RQ3 What devices have been used in these VR usability studies employing HE across different domains?

Table 6: Devices Used Across Domains.

Domain and Studies	Devices	Devices Categories	Total Studies
Core HCI & technology [8, 26, 27, 28, 29, 31, 33, 34, 36, 38, 39, 40, 43, 44, 46, 47, 48]	Desktop (3), Desktop + Data glove (1), Desktop + head-tracking and a wand (1), notebook + HTC Vive Pro HMD (1), VR headset (6), Desktop + Controller (1), Glove-such as device (1), mirror-based VR display (1), Immersive CUBE (1), HTC Vive + Oculus Rift (1)	Desktop and its “variants” (7); Wearable (8); Glove as device (1), Immersive CUBE (1)	17
Games [53, 7, 54, 55, 18, 37, 56, 57].	Desktop (2); Oculus Rift (4); Google Cardboard (1); HMD (1);	Desktop (2); Wearable (6)	8
Education [58, 59, 60, 61, 62, 80, 81, 82, 89].	Oculus Rift (1); Desktop (2); HTC Vive (1); Oculus Go (1); Smartphones/VR BOX (1); Oculus Quest (2); OpenSimulator VR (1)	Wearable (7); Desktop (2)	9
Tourism & cultural Heritage [69, 70, 71, 72, 85, 86].	Workbench custom-made unified laser pointer that emits both infrared (IR) and visible light (RED) + Ceiling-mounted IR camera (1); Smartphone (1); HMD (1); Google Cardboard (1); Oculus Quest 2/360° Cameras/LiDAR (1); Meta Quest Pro (1)	Wearable (4); Smartphone (1); Workbench (1)	6
Healthcare, Psychology & Treatment [51, 52, 73, 87, 88].	HMD (1); HTC Vive (1); not specified (1); Meta Quest 3 (1); Oculus HMD + Desktop UI (1)	Wearable (3); Hybrid Wearable/Desktop (1); Not specified (1)	5
Training & Industrial Applications [11, 64, 65, 66, 68, 83, 84]	Desktop + stereoscopic visualization + Nintendo Wii Remote (1); Desktop (1); HMD (2); Oculus Quest2 (1); HTC Vive Pro + ROS (1); HTC VIVE Pro/Meta Quest 3/VIVE XR Elite + Desktop (1)	Wearable (3); Desktop (2); Hybrid Wearable/Desktop (2)	7
Marine Archaeology [45]	Desktop + 3D Glasses, X-box controller and VR CAVE (1)	Hybrid system (1)	1

Table 6 details the intricate relationships between the research domain and the technologies used, including VR systems. We have categorised immersive devices (e.g., Head tracking, HTC Vive Pro, VR headset, VR display, Oculus Rift, Google Cardboard, Oculus Go, HMD, Oculus Quest 2, Meta Quest 3,

Meta Quest Pro, and VIVE XR Elite) as 'Wearables', configurations relying on traditional monitors as 'Desktops', and specialised finger-tracking tools as 'Glove-like devices'.

The Core HCI & Technology domain comprises 17 studies and uses various devices, including desktop computers, wearables, controllers, and specialised equipment such as data gloves and mirror-based displays. Within the Games domain, eight studies were included. Desktop computers account for a large portion of the hardware here, whilst iconic wearables, such as the Oculus Rift, Google Cardboard, and versatile HMDs, are widely utilised. This device choice focuses on UX within diverse game genres, ensuring players navigate VEs with smooth interaction and immersive engagement.

The Education domain includes 9 studies that indicate a significant interest in VR's educational applications, each exploring VR's transformative potential in learning environments. The Oculus Rift and Quest series headsets, as well as desktop computers, have emerged as preferred tools, providing a robust platform for delivering interactive lessons, fostering collaboration, and enhancing knowledge retention.

The Tourism and Cultural Heritage domain showcases the growing fascination with virtual travel experiences. Devices range from specialised items, such as workbenches and laser pointers, to ubiquitous smartphones and wearables, such as the Meta Quest Pro. This variety reflects the multifaceted nature of VR tourism evaluations, encompassing virtual tours, cultural exploration, and even the simulation of real-world travel scenarios.

Healthcare and Psychology domains contributed five studies, utilising wearables and advanced hybrid interfaces to explore the impact of VR on behaviour, emotions, and clinical treatments. Meanwhile, the Training and Industrial domain boasted seven studies emphasising highly practical applications. Wearables and advanced standalone headsets are crucial tools in these domains, enabling researchers to monitor spatial behaviours, track body movements, and safely simulate high-risk environments.

Diverse domains, such as Marine Archaeology, contribute to a single study, showcasing the vast reach and versatility of VR applications. The devices employed in these studies reflect their unique foci, highlighting the ongoing exploration of VR's potential across diverse domains of human endeavour.

Recent studies (2024–2025) highlight a distinct shift toward advanced standalone and mixed-reality-capable headsets, such as the Meta Quest 3, Meta Quest Pro, and VIVE XR Elite [84, 86, 87]. Furthermore, domains such as Healthcare [88], Industrial Robotics [83], and Additive Manufacturing [84] increasingly utilise "Hybrid" or "Asymmetrical" systems. In these setups, primary users wear wearables while supervisors or therapists use traditional Desktop interfaces to monitor the simulation. This device choice ensures that users navigating the VE experience immersive engagement while operators maintain system control and avoid VR-induced fatigue.

The analysis of devices employed in VR research (RQ3) provides valuable insights into the evolving technological landscape. The dominance of wearables reflects their increasing accessibility and affordability for researchers across diverse domains. However, it is essential to acknowledge the limitations of relying solely on wearables, as certain domains may benefit from incorporating other devices, such as desktops or specialised equipment, into their evaluation setups (Table 6).

To better visualise these hardware trends, Figure 4 illustrates the distribution of device categories across all evaluated domains. The data clearly demonstrates the pervasive adoption of wearables as the industry standard for immersion. However, the chart also visualises a critical emerging trend: the strategic use of Hybrid systems (combining wearable for the end-user with 2D Desktop interfaces for the operator) specifically in high-stakes, complex domains such as Training & Industrial, Healthcare, and Marine Archaeology. This shift indicates that while wearables drive the user experience, supplementary desktop interfaces remain necessary for system control and evaluation.

Furthermore, the discussion on challenges encountered with general-purpose heuristics (RQ4a) emphasises the need to consider device-specific factors during VR usability evaluations. Studies employing wearables, for example, might encounter unique challenges related to user comfort, motion-tracking accuracy, and potential limitations in specific interaction tasks. Tailoring evaluation methods to account for these device-specific considerations is crucial to ensure reliable and insightful results.

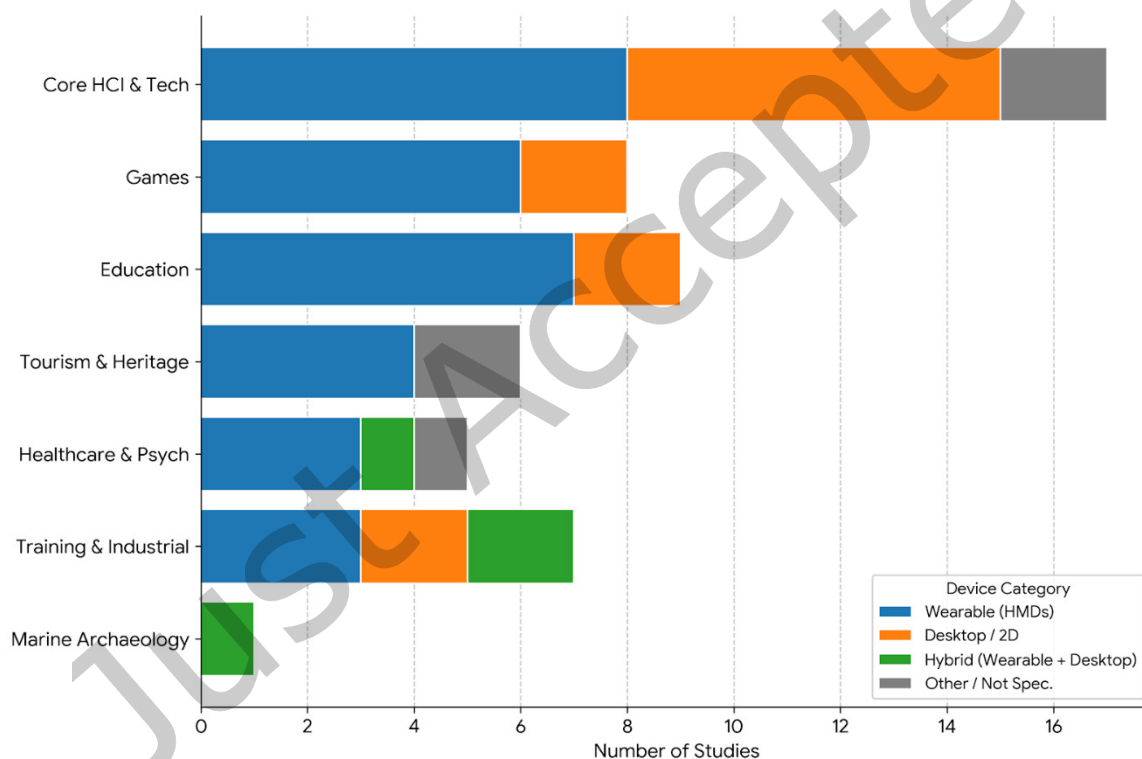


Figure 4: Distribution of Device Categories across Domain

RQ4a **What challenges have been encountered when applying general-purpose heuristics to VR evaluations compared to VR-specific heuristics?**

Applying familiar tools to a new domain may present unexpected challenges. This is particularly true when using general-purpose heuristics to evaluate VR applications. Studies such as [26] highlight the limitations of traditional approaches, including Nielsen's heuristics [6]. VR's immersive nature often introduces novel challenges, such as motion sickness and discomfort related to 3D spatial interactions [74]. As [26] emphasises, VR evaluations require a heightened focus on representation fidelity, encompassing visual elements and other factors influencing usability.

Researchers in [39] offer another compelling example. Their study applied Nielsen's [30] heuristics to evaluate dynamic guidance systems in VR transit hubs. They discovered a critical issue that traditional methods could not detect: cybersickness induced by the immersive environment. Recent studies further amplify these concerns in multi-user and high-stakes environments. For instance, collaborative VR introduces significant network latency, disrupting communication and inducing UI lag [81, 87]. In research on industrial and cultural applications, merging 3D reconstructions with 360-degree video can cause visual clashes and "eye-level" illusions [85], while rendering complex 3D annotations in mid-air often leads to disorienting "depth errors" [84]. A study evaluating cultural heritage simulations found that while the interface scored high on traditional heuristics, the spatial reconstructions felt emotionally "empty" without deep narrative integration [86]. In mental health settings, researchers discovered the "therapist/operator blindspot", in which operators wearing VR headsets missed critical physical patient cues (e.g., sweating or skin flushing) that traditional heuristics do not account for [88].

Moreover, [48] took a different approach, combining Nielsen's [6] heuristics with principles from [48]. Their research revealed specific challenges for participants unfamiliar with computers or application metaphors. This finding underlines the importance of developing VR-specific heuristics tailored to these unique scenarios. In educational settings, traditional heuristics similarly fail to measure cognitive load or the integration of learning theories within immersive 3D spaces [89].

This underscores the need to carefully explore the suitability of immersive VR for HE across diverse domains. Furthermore, general heuristics fail to capture the nuances of emotional and physiological. The discussion on challenges encountered with general-purpose heuristics (RQ4a) highlights their limitations in capturing the unique characteristics of VR environments. These limitations often manifest as an inability to address key aspects of the VR user experience, including immersion, presence, embodiment, and flow. Additionally, studies such as [40] emphasise the potential for cybersickness and discomfort induced by VR's immersive nature, which traditional heuristics may not adequately capture.

RQ4b: What solutions have been proposed by authors in relevant studies to address the challenges identified in RQ4a?

Several studies offer valuable insights into overcoming the limitations of general-purpose heuristics in VR evaluation. Researchers in [28] advocate for a combined approach. While acknowledging the usefulness of Nielsen's [6] heuristics, they suggest further analysis to develop dedicated VR-specific guidelines, especially for function-oriented applications. Slone [29] echoes this sentiment, highlighting the importance of accurate evaluator training to avoid false positives when blending heuristics, such

as Nielsen's [6], with checklists such as the Xerox usability guidelines. Domain-specific expertise becomes crucial for effectively detecting and addressing usability flaws before they reach users.

The challenges extend beyond traditional usability concerns. Researchers in [18] encountered limitations when using Nielsen and Molich's heuristics [30] to evaluate VR games, due to complexities such as multiple HMDs and natural-language interactions. Their case emphasises the need for dedicated VR game heuristics. Similarly, researchers in [58] used the heuristics from [30] to evaluate virtual clinical simulations and encountered disorienting and nauseating effects not covered by the existing framework.

To bridge these historical gaps, recent studies have proposed innovative technological and UI solutions to address VR's unique challenges. To combat collaborative latency and motion sickness, researchers implemented teleportation locomotion and high-frequency (0.1s) data updates [81]. To overcome depth errors in 3D spatial drawing, [84] introduced 2D "slicing planes" for annotations and a Desktop Spectator Mode to lower the hardware learning curve for industrial experts. UI clutter in complex robotic systems [83] was mitigated using dropdown lists and universally recognised, culturally neutral icons.

In healthcare and education, reading heavy text panels in VR can cause severe eye strain. Replacing these panels with Generative AI-powered Natural Language Processing (NLP) avatars [82, 87] significantly reduced cognitive load and fostered positive emotional responses. To address the emotional emptiness in heritage sites, [86] grounded tasks in rich narrative storylines. Finally, to address the "operator blindspot" in clinical therapy, [88] proposed a dual-interface design that combines a 3D VR interface to build trust with the patient and a 2D desktop interface for the therapist to maintain control. Finally, to address the pedagogical shortcomings of general heuristics, [89] developed the WVREA framework, which provides a validated set of heuristics specifically targeting immersion, cognitive load, and educational outcomes in wearable VR apps.

With the emergence of AI-assisted tools in recent VR evaluations [82, 86], an important avenue opens for leveraging artificial intelligence to support or partially automate heuristic evaluation itself. While traditional heuristic evaluation guidelines rely heavily on 3–5 expert evaluators [6], the growing scarcity of domain-specific VR experts, particularly in education, healthcare, and industry, makes this standard increasingly difficult to follow in practice.

AI technologies, such as large language models for automated severity rating, computer vision for spatial issue detection, or generative agents for simulating diverse user perspectives [87], could serve as scalable assistants to human experts. In such a hybrid model, experts would retain their critical role as validators and final arbiters, ensuring contextual nuance and ethical judgment, while AI handles repetitive pattern recognition and initial triage. This shift could significantly reduce evaluation time and costs, democratise access to rigorous HE in resource-constrained settings, and enable broader participation by non-expert stakeholders (e.g., educators or clinicians). Future research should empirically test the reliability and validity of such AI-augmented heuristic protocols, particularly in higher education applications where domain expertise is often limited but pedagogical impact is high.

While some studies encountered limitations, others offered examples of successful adaptations. Researchers [66, 71] successfully employed combinations of VR heuristics for crisis management and educational apps. For example, Researchers in [66] demonstrated the effectiveness of combining

twelve heuristics [7] in a crisis management training simulator. They identified the need for a multi-heuristic approach for comprehensive evaluations. Furthermore, [71] also successfully employed the heuristics of [7] for educational apps, leading to a recommended redesign focused on pedagogical benefits, diverse assessment types, and teacher analytical tools. Furthermore, [73] successfully applied a 20-item HE questionnaire to assess eating disorder treatment apps, focusing on infrastructural, technological, and financial aspects, without encountering significant VR environment-related issues.

Researchers in [27] faced challenges in the HE of adaptive VR due to the complexities of the technology itself, necessitating further refinement of their lightweight rate adaptation heuristics, whilst [65] also encountered limitations in their Design-based Research (DBR) Heuristic for VR training programs, highlighting the need for ongoing improvements in VR-specific evaluation methodologies.

Table 7: Heuristic Distribution by Year, Domain, and Device

Year	Total	General Heuristics (Count)	Specific / Mixed Heuristics (Count)	Heuristics Not Specified (Count)
2004	2	Core HCI (1)	Games (1)	
2009	2		Games (1); Core HCI (1)	
2010	1		Training (1)	
2011	3	Education (1); Games (1)	Core HCI (1)	
2012	2		Training (1); Core HCI (1)	
2013	2	Core HCI (1)	Core HCI (1)	
2014	4	Training (1)	Education (1); Core HCI (1); Tourism (1)	
2015	1		Marine Arch. (1)	
2016	4	Tourism (1); Core HCI (1)	Tourism (1); Games (1)	
2017	2		Core HCI (1); Games (1)	
2018	1		Games (1)	
2019	5	Core HCI (1)	Tourism (1); Games (1)	Psychology (1); Technology (1)
2020	4	Education (1)	Education (1)	Core HCI (1); Core HCI (1)
2021	5	Core HCI (1); Education (1)	Psychology (1); Healthcare (1); Games (1)	
2022	2		Core HCI (1); Training (1)	
2023	3		Core HCI (1); Core HCI (1); Training (1)	
2024	5	Healthcare (1); Training/Ind. (1)	Education (1); Training/Industry. (1) Tourism (1)	
2025	5		Education (3); Tourism (1); Healthcare (1)	
TOTAL	53	13	36	4

However, modern researchers increasingly recognise that HE alone is insufficient. Recent evaluations have aggressively paired heuristics with mixed-methods approaches, such as the System Usability Scale (SUS) [87], Concurrent Think-Aloud (CTA) protocols [86], and Data Mining [82] to capture the holistic user experience. In conclusion, the VR usability evaluation presents challenges and opportunities. While general-purpose heuristics offer a valuable starting point, their limitations become apparent when navigating the unique complexities of the virtual environment. Embracing a combination of existing frameworks, tailoring them to specific domains, and developing dedicated VR-specific heuristics are crucial to ensure comprehensive and effective usability evaluations for this growing technology. Through ongoing research and adaptation, we can chart a course to optimise UX and unlock the full potential of VR across diverse applications.

RQ5 Based on the identified trends in existing VR HE studies, what research areas hold the most potential to contribute effectively to improving usability and UX in VR systems?

Figure 5 reveals the temporal trajectory of VR heuristic evaluation research from 2004 to 2025. Rather than interpreting individual year-to-year fluctuations—which are sensitive to small sample sizes—the overall pattern reveals three distinct phases of intellectual development, each characterised by a progressively deeper engagement with VR's unique properties.

Phase 1: Adaptation (2004–2015). Early studies ($n=17$) focused on a foundational question: can traditional 2D heuristics be stretched to cover basic VR tasks? For example, early studies [26, 33] applied Nielsen's principles to desktop-VR and CAVE environments, treating VR as an extension of conventional graphical interfaces rather than a distinct medium. These studies identified initial mismatches—particularly around 3D navigation and spatial presence—but rarely proposed systematic alternatives.

Phase 2: Specialisation (2016–2021). The second phase ($n=21$) saw the emergence of dedicated VR-specific heuristic sets, most notably Sutcliffe and Gault's 12 VR Heuristics [7] and experimental virtual world heuristics [37]. However, these frameworks remained largely domain-agnostic, designed to capture generic immersive properties (presence, navigation, embodiment) without addressing the contextual demands of specific application areas. The peak activity in 2019 and 2021 reflects this period of framework proliferation.

Phase 3: Contextual Integration (2022–2025). The most recent phase ($n=15$) marks a decisive shift from standalone checklists to heuristics embedded within broader socio-technical evaluation systems. Three convergent strategies define this phase. First, mixed-method instrumentation: heuristics are paired with quantitative scales (SUS, CTA) and physiological metrics to capture dimensions—such as cognitive load and emotional resonance—that checklists alone cannot detect [82, 86, 87]. Second, AI-mediated evaluation: generative AI agents and NLP avatars are being integrated as both evaluation targets and assistive tools, requiring heuristics that assess dynamic, context-aware interactions rather than static interface elements [82, 87]. Third, asymmetrical interface design: in high-stakes domains (healthcare, industrial training), heuristics must now evaluate dual-interface systems where VR users interact with non-VR operators, introducing new usability concerns such as the "operator blindspot" [88] and collaborative latency [81, 84]. Together, these strategies reflect a maturation from "can HE

work in VR?" to "how must HE be restructured to account for VR as a collaborative, embodied, and computationally augmented medium?"

This phased progression suggests that the field is converging on a new evaluation paradigm: one that treats VR usability not as a property of individual interfaces, but as an emergent property of distributed human-AI-technical systems operating within specific domain constraints.

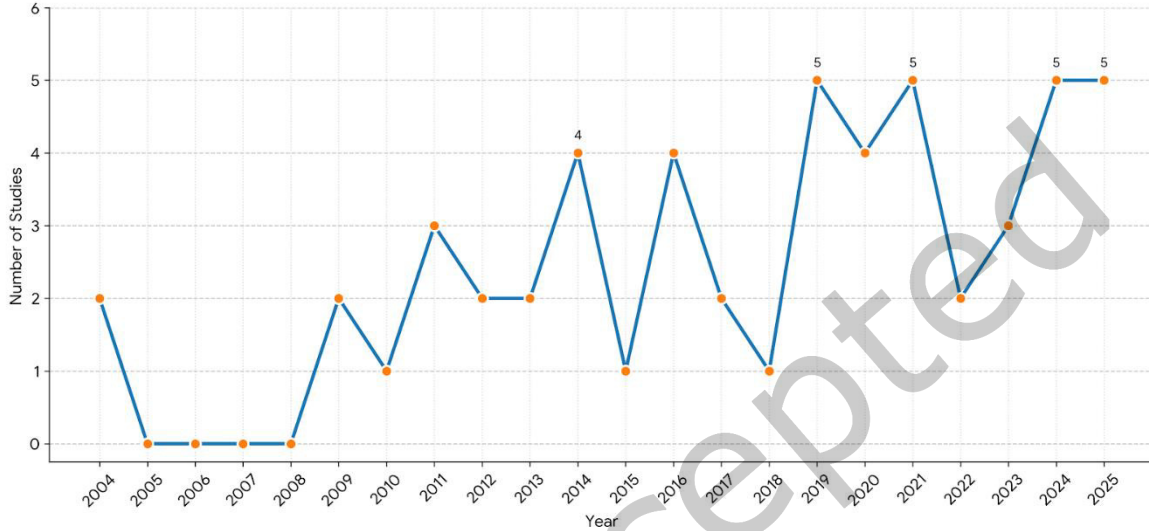


Figure 5: Trends of HE in VR systems between 2004 and 2025

Table 7 summarises different heuristics applications in VR usability research. Historically, studies employing general heuristics were most prevalent in Core HCI and Technology. However, in recent years, there has been a marked shift toward Specific and Mixed-Method heuristics across all domains.

Notably, many earlier studies (Figure 6) focused solely on wearables and used general heuristics, suggesting a need for further exploration to better understand the phenomenon.

Specific heuristics show a much more diverse pattern (Figure 7). Core HCI & Technology utilised specific heuristics across various devices, showcasing this domain's dedication to refining fundamental UX. Education and Gaming closely follow, emphasising wearable-centric evaluations to capture immersion and cognitive load [89].

A deeper analysis reveals a gap within other domains, and their historical reliance on existing frameworks highlights the absence of dedicated VR heuristics. Despite Sutcliffe and Gault's [7] call for a revision of the VR heuristic nearly two decades ago, this lack of domain-specific tools remains a crucial area for future research. Moreover, an interesting pattern emerged: between 2015 and 2025, the vast majority of studies (over 35) utilised Wearables or Hybrid (Wearable + Desktop) systems, whilst studies relying strictly on standalone desktops plummeted. This indicates that wearable technologies are affordable and accessible, serving as the definitive standard for modern researchers

across all domains. Based on the identified trends and gaps in existing research, several promising avenues for future exploration emerge:

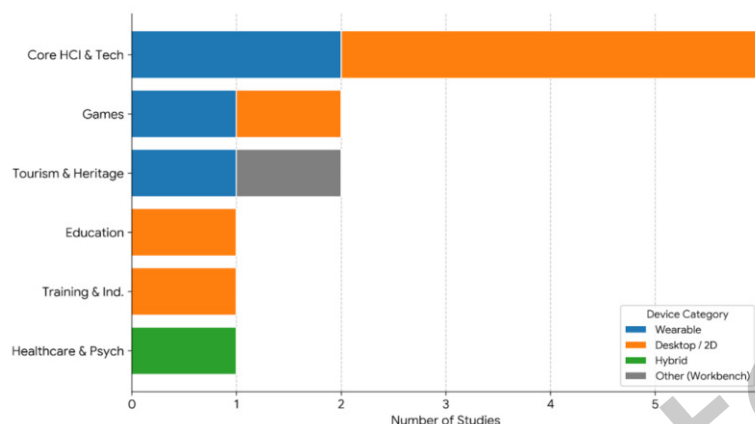


Figure 6: Research Domain and Device Types for General Heuristic (N=13)

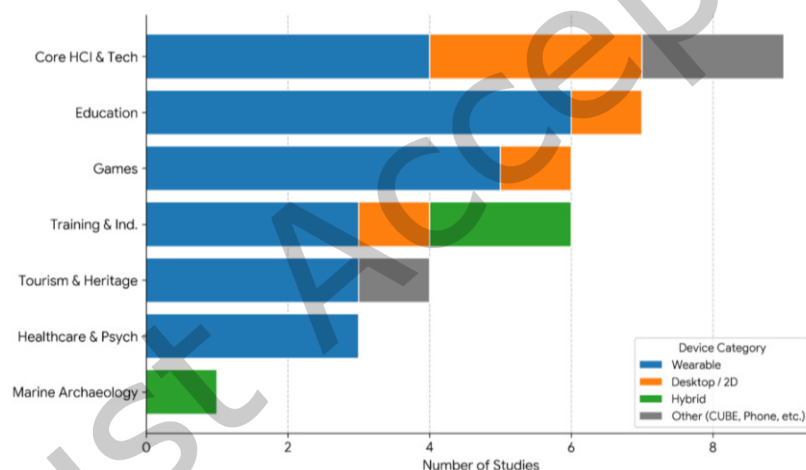


Figure 7: Research Domain and Device Types for Specific/Mixed Heuristics

1. Prioritising development of domain-specific and Multi-User VR heuristics: This is particularly crucial for collaborative environments (Education, Telemedicine, Industrial Digital Twins) which rely heavily on network synchronisation and co-presence [81, 84, 87]. General heuristics fail to measure the usability of social interaction and latency.
2. Integration of Generative AI and LLMs: Research evaluating the usability of AI-driven Virtual Humans and NLP avatars is rapidly expanding [82, 87]. Future UX studies will need to evaluate how effectively AI can provide real-time guidance, dynamic narratives, and context-aware VR textures to reduce cognitive load.

3. Cross-Reality (XR) and Asymmetrical Interfaces: As VR moves into high-stakes industrial and healthcare settings, future research must evaluate dual-interface systems in which users in VR interact seamlessly with operators or supervisors using 2D desktop or Mixed Reality (MR) interfaces [84, 88].
4. Refining Mixed-Method Evaluations: Moving forward, HE must be continuously paired with physiological data (e.g., biometric wristbands for stress monitoring) [88] and standardised psychological metrics (SUS, CTA) [86, 87] to capture the holistic user experience that heuristic checklists alone cannot detect.

Table 8: Summary of key implications emerging from the systematic review trends and gaps (N=53)

Gap/Trend Identified from the Review	Real-World Implication	Relevant Domains & Example Studies	Suggested Future Action / Research Direction
Fragmentation & low reuse of heuristics (most used <2 times)	Increased development time/cost for VR applications due to repeated ad-hoc heuristic creation.	Education, Training, Healthcare (e.g., [31, 37, 58, 60])	Longitudinal validation of reusable frameworks such as WVREA [89] across diverse HE contexts.
Emergence of AI-assisted elements in VR evaluation	Scalable automation of issue detection, real-time feedback, and reduced cognitive load for users/evaluators.	Teacher Education, Telemedicine, Cultural Heritage (e.g., [82, 86, 87])	Develop AI-augmented heuristic guidelines (e.g., automated severity scoring + bias detection in avatars).
Rise of hybrid/asymmetrical interfaces	Mitigates evaluator/operator fatigue, lowers learning curves for novices, and enables safe oversight in high-stakes settings.	Mental Health, Nuclear Robotics, Additive Manufacturing (e.g., [83, 84, 88])	Standardise asymmetrical evaluation protocols for multi-user VR in collaborative and clinical/industrial use.
Underrepresentation of cultural/contextual adaptation	Reduced accessibility/adoption in non-Western or diverse global HE contexts due to Western-centric heuristics.	Tourism & Cultural Heritage, Education (e.g., [85, 86, 89])	Conduct cross-cultural empirical validations and adaptations of existing heuristics (e.g., WVREA extensions).

Integrative Synthesis of Findings

Collectively, the findings of RQ1–RQ4 reveal that VR heuristic evaluation remains in a transitional phase. While RQ1 demonstrates a fragmented landscape dominated by custom and mixed-method approaches, RQ2 and RQ3 show clear domain expansion (especially into Education, Healthcare, and Industrial Training) and a hardware shift toward wearables and hybrid/asymmetrical systems. However, RQ4a exposes persistent limitations: traditional/general heuristics systematically fail to capture VR's unique spatial, cognitive, and social complexities (e.g., simulator sickness, multi-user latency, operator blind spot). RQ4b highlights early progress through AI integration and hybrid designs, yet these solutions remain scattered and under-validated. The overarching critical gap is the lack of scalable, standardised, and domain-sensitive heuristic frameworks to support VR's rapid maturation in high-stakes contexts. Without such frameworks, development costs remain high, accessibility is limited in resource-constrained regions, and transformative potential in higher education and clinical training is constrained. Addressing this gap through synthesis, validation, and cross-domain adaptation represents the most pressing direction for the field in high-stakes contexts such as higher education and clinical training, where usability directly affects learning outcomes and patient safety. These interconnected trends and their real-world implications are summarised in Table 8.

Limitations of the Study

While this systematic review provides a comprehensive analysis of peer-reviewed literature, certain limitations must be acknowledged. First, to ensure a high standard of methodological rigour and external validation, our search protocol was strictly limited to four major electronic databases (SCOPUS, IEEE Xplore, ACM Digital Library, and Web of Science). Consequently, this review excluded grey literature, such as unpublished master's theses, doctoral dissertations, and industry white papers. This exclusion criterion was purposefully applied to avoid data duplication (double-counting), as high-quality graduate research in HCI is typically serialised and published in peer-reviewed venues shortly after completion (a trajectory evident in recent frameworks such as WVREA [89]). While prioritising peer-reviewed literature ensures that the analysed heuristics have undergone rigorous academic scrutiny, we acknowledge that excluding grey literature may omit early-stage industry work or student-led projects containing novel, albeit unvalidated, VR usability heuristics. Future systematic reviews could benefit from incorporating dedicated grey literature searches (e.g., via ProQuest Dissertations & Theses) to capture these ongoing contributions and further mitigate potential publication bias. Second, our search was restricted to English-language articles, which may have excluded relevant heuristic frameworks developed and published in other languages.

Finally, although the search protocol extended through February 26, 2026, no eligible publications from 2026 were identified during screening. This reflects the timing of peer-review cycles and publication lags for emerging VR research. Consequently, the effective temporal scope of the included evidence is 2004–2025. Moreover, the VR hardware landscape (e.g., eye-tracking, mixed-reality pass-through) is evolving rapidly; therefore, some very recent hardware-specific heuristics may still be in peer review. Readers should note that very recent advancements, such as post-February 2026 developments in eye-tracking heuristics (Vision Pro ecosystem) or generative AI for adaptive VR interfaces, may introduce additional frameworks not captured here. Future replications or rapid updates to this review would be valuable as the field continues to evolve.

6 CONCLUSION

This systematic literature review comprehensively analysed heuristic evaluation methodologies applied to VR systems from 2004 to 2025. Across the 53 documented studies, researchers employed a highly fragmented landscape of frameworks, which this review categorised into four distinct approaches: the use of standalone VR-Specific & Custom Heuristics ($n=22$, 41.5%), Combinations & Mixed-Methods ($n=14$, 26.4%), Traditional/General Heuristics ($n=13$, 24.5%), and evaluations where the framework was not specified ($n=4$, 7.6%). The research landscape was systematically mapped across seven distinct application domains: Core HCI & Technology, Education, Games, Training & Industrial, Tourism & Cultural Heritage, Healthcare & Psychology, and Marine Archaeology. Regarding hardware, while evaluations heavily favoured immersive Wearables (e.g., HTC Vive, Oculus, Meta Quest), there is an emerging, strategic use of Hybrid/Specialised systems (combining HMDs with 2D desktop monitoring) in complex industrial and clinical settings.

Notably, this review confirms that general heuristics (such as Nielsen's traditional principles) often fail to detect usability issues specific to immersive 3D environments. Traditional frameworks routinely overlook critical physical, spatial, and hedonic qualities such as immersion, embodiment, spatial UI clutter, and simulator sickness [86]. Furthermore, as VR scales into collaborative and high-stakes

sectors, general heuristics cannot account for complex modern phenomena such as multi-user network latency [81], the heavy cognitive load of inspecting and recording issues within VR [11], or the "operator blindspot" (where therapists or trainers lose physical user cues when wearing HMDs) [88].

In conclusion, the landscape of VR usability research presents a wealth of opportunities to advance user experience. By systematically addressing domain-specific research gaps, transitioning away from traditional 2D heuristics toward validated mixed-methods, and venturing into new domains with tailored frameworks, future research can ensure that VR systems not only captivate audiences but also deliver seamless, safe, and effective experiences that unlock the full potential of this transformative technology.

ACKNOWLEDGMENTS

The authors would like to acknowledge Universiti Malaysia Sarawak (UNIMAS) for the support provided towards this publication.

We acknowledge the grant from the Universiti Malaysia Sarawak. [UNI/F04/GRADUATES/85191/2022]

REFERENCES

- [1] A. Abulfaraj and A. Steele. 2020. Coherent heuristic evaluation (CoHE): Toward increasing the effectiveness of heuristic evaluation for novice evaluators. In *International Conference on Human-Computer Interaction*, 2020. Cham: Springer International Publishing, 3–20.
- [2] Estela Aparecida Oliveira Vieira, Aleph Campos da Silveira, and Ronei Ximenes Martins. 2019. Heuristic evaluation on usability of educational games: A systematic review. *Informatics in Education* 18, 2 (2019), 427–442. <https://doi.org/10.15388/infedu.2019.20>
- [3] Daniela Quiñones and Cristian Rusu. 2019. Applying a methodology to develop user eXperience heuristics. *Comput Stand Interfaces* 66, (October 2019), 103345. <https://doi.org/10.1016/j.csi.2019.04.004>
- [4] Takudzwa Mujuru and Christian Lopez. 2021. Creating virtual reality teaching modules for low-cost headsets. In *Volume 2: 41st Computers and Information in Engineering Conference (CIE)*, August 17, 2021. American Society of Mechanical Engineers. <https://doi.org/10.1115/DETC2021-72084>
- [5] Brad Thompson. 2020. Work-in-progress—learning machines: Useful intelligent tutoring within immersive STEM VLEs. In *2020 6th International Conference of the Immersive Learning Research Network (ILRN)*, June 2020. IEEE, 247–250. <https://doi.org/10.23919/iLRN47897.2020.9155146>
- [6] Jakob Nielsen. 1994. Usability inspection methods. In *Conference companion on Human Factors in Computing Systems*, 1994. 413–414.
- [7] Alistair Sutcliffe and Brian Gault. 2004. Heuristic evaluation of virtual reality applications. In *Interacting with Computers*, 2004. Elsevier, 831–849. <https://doi.org/10.1016/j.intcom.2004.05.001>
- [8] Rabia Murtza, Stephen Monroe, and Robert J. Youmans. 2017. Heuristic evaluation for virtual reality systems. In *Proceedings of the Human Factors and Ergonomics Society*, 2017. Human Factors and Ergonomics Society Inc., 2067–2071. <https://doi.org/10.1177/1541931213602000>
- [9] Joseph L Gabbard, H Rex Hartson, and Sallie M Henry. 1997. A taxonomy of usability characteristics in virtual environments. Virginia Tech.
- [10] Andrej Somrak, Matevž Pogačnik, and Jože Guna. 2021. Suitability and comparison of questionnaires assessing virtual reality-induced symptoms and effects, and user experience in virtual environments. *Sensors* 21, 4 (February 2021), 1185. <https://doi.org/10.3390/s21041185>
- [11] Xiaolei Guo, Kushal Kumar Nerella, Jiahui Dong, Zhenyu Qian, and Yingjie Chen. 2024. Understanding pitfalls and opportunities of applying heuristic evaluation methods to VR training systems: An empirical study. *International Journal of Human-Computer Interaction*, 2023. <https://doi.org/10.1080/10447318.2022.2161238>
- [12] David Checa and Andres Bustillo. 2020. A review of immersive virtual reality serious games to enhance learning and training. *Multimed Tools Appl* 79, 9–10 (March 2020), 5501–5527. <https://doi.org/10.1007/s11042-019-08348-9>

- [13] Dorota Kamińska, Tomasz Sapiński, Sławomir Wiak, Toomas Tikk, Rain Eric Haamer, Egils Avots, Ahmed Helmi, Cagri Ozcinar, and Gholamreza Anbarjafari. 2019. Virtual reality and its applications in education: Survey. *Information (Switzerland)* 10, 10 (October 2019). <https://doi.org/10.3390/info10100318>
- [14] S. Sriworapong, A. Pyae, A. Thirasawasd, and W. Keereewan. 2022. Investigating students' engagement, enjoyment, and sociability in virtual reality-based systems: A comparative usability study of spatial. io, Gather. town, and Zoom. In *International Conference on Well-Being in the Information Society*, 2022. Cham: Springer International Publishing, 140–157. https://doi.org/10.1007/978-3-031-14832-3_11
- [15] Jenny Morales, Héctor Cornide-Reyes, Pedro O. Rossel, Paula Sáez and Fabián Silva-Aravena. 2023. Virtual reality, augmented reality and metaverse: Customer experience approach and user experience evaluation methods. *Literature review*. 554–566. https://doi.org/10.1007/978-3-031-35915-6_40
- [16] Hashem A. Almusawi, Christopher M. Durugbo, and Afaf M. Bugawa. 2021. Wearable technology in education: A systematic review. *IEEE Transactions on Learning Technologies* 14, 4 (August 2021), 540–554. <https://doi.org/10.1109/TLT.2021.3107459>
- [17] Hossain Samar Qorbani, Ali Arya, Nuke Nowlan, and Maryam Abdinejad. 2021. ScienceVR: A virtual reality framework for STEM education, simulation and assessment. In *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, November 2021. IEEE, 267–275. <https://doi.org/10.1109/AIVR52153.2021.00060>
- [18] Daniel Paes and Javier Irizarry. 2018. A usability study of an immersive virtual reality platform for building design review: Considerations on human factors and user interface. In *Construction Research Congress 2018: Construction Information Technology - Selected Papers from the Construction Research Congress 2018*, 2018. American Society of Civil Engineers (ASCE), 419–428. <https://doi.org/10.1061/9780784481264.041>
- [19] Francisco Rebelo, Paulo Noriega, Emília Duarte, and Marcelo Soares. 2012. Using virtual reality to assess user experience. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54, 6 (December 2012), 964–982. <https://doi.org/10.1177/0018720812465006>
- [20] Niels Hinricher, Simon König, Chris Schröer, and Claus Backhaus. 2023. Effects of virtual reality and test environment on user experience, usability, and mental workload in the evaluation of a blood pressure monitor. *Front Virtual Real* 4, (June 2023). <https://doi.org/10.3389/frvir.2023.1151190>
- [21] A. A. D. Luna. 2022. *Introduction to virtual reality*. Arcler Press.
- [22] Ilsun Rhiu, Yong Min Kim, Wonjoon Kim, and Myung Hwan Yun. 2020. The evaluation of user experience of a human walking and a driving simulation in the virtual reality. *Int J Ind Ergon* 79, (September 2020), 103002. <https://doi.org/10.1016/j.ergon.2020.103002>
- [23] Rodolfo Inostroza, Cristian Rusu, Silvana Roncagliolo, Virginica Rusu, and César A. Collazos. 2016. Developing SMASH: A set of SMARtphone's uSability Heuristics. *Comput Stand Interfaces* 43, (January 2016), 40–52. <https://doi.org/10.1016/j.csi.2015.08.007>
- [24] Barbara Kitchenham. 2004. *Procedures for performing systematic reviews*.
- [25] David Moher. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann Intern Med* 151, 4 (August 2009), 264. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- [26] Xuesong Zhang and Adalberto L. Simeone. 2021. Using heuristic evaluation in immersive virtual reality evaluation. In *20th International Conference on Mobile and Ubiquitous Multimedia*, May 12, 2021. ACM, New York, NY, USA, 223–225. <https://doi.org/10.1145/3490632.3497863>
- [27] Jeroen Van Der Hooft, Maria Torres Vega, Stefano Petrangeli, Tim Wauters, and Filip De Turck. 2019. Tile-based adaptive streaming for virtual reality video. *ACM Transactions on Multimedia Computing, Communications and Applications* 15, 4 (December 2019). <https://doi.org/10.1145/3362101>
- [28] Wenting Wang, Jinghui Cheng, and Jin L.C. Guo. 2019. Usability of virtual reality application through the lens of the user community. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, May 02, 2019. ACM, New York, NY, USA, 1–6. . <https://doi.org/10.1145/3290607.3312816>
- [29] Debra J Slone. 2009. A methodology for measuring usability evaluation skills using the constructivist theory and the second life virtual world. *J Usability Stud* 4, 4 (2009), 178–188.
- [30] Jacob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, March 1990. 249–256.
- [31] Pascal Jansen, Julian Britten, Alexander Häusele, Thilo Segschneider, Mark Colley, and Enrico Rukzio. 2023. AutoVis: Enabling mixed-immersive analysis of automotive user interface interaction studies. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, April 19, 2023. Association for Computing Machinery, 1–23. <https://doi.org/10.1145/3544548.3580760>

- [32] Dan R. Olsen. 2007. Evaluating user interface systems research. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, October 07, 2007. ACM, New York, NY, USA, 251–258. <https://doi.org/10.1145/1294211.1294256>
- [33] Jonatan Martínez, Arturo S. García, José P. Molina, Diego Martínez, and Pascual González. 2013. An empirical evaluation of different haptic feedback for shape and texture recognition. *Visual Computer* 29, 2 (February 2013), 111–121. <https://doi.org/10.1007/s00371-012-0716-x>
- [34] Muhammad Shafie Abdul Latif and Setyawan Widyarto. 2004. The crowd simulation for interactive virtual environments. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry - VRCAI '04*, 2004. ACM Press, New York, New York, USA, 278. <https://doi.org/10.1145/1044588.1044647>
- [35] Srikanth Bandi and Daniel Thalmann. 2000. Pathfinding for human motion in virtual environments. *Computational Geometry* 15, 1–3 (February 2000), 103–127. [https://doi.org/10.1016/S0925-7721\(99\)00046-2](https://doi.org/10.1016/S0925-7721(99)00046-2)
- [36] Jesús David Cardona, Jenny Medeiros, and Andrés Solano. 2016. Design and usability evaluation of virtual environments created on the 3DGEN platform. In *ACM International Conference Proceeding Series*, September 13, 2016. Association for Computing Machinery. <https://doi.org/10.1145/2998626.2998647>
- [37] Cristian Rusu, Roberto Muñoz, Silvana Roncagliolo, Sebastian Rudloff, Virginica Rusu, and Arturo Figueroa. 2011. Usability heuristics for virtual worlds. In *Proceedings of the Third International Conference on Advances in Future Internet*, August 2011. 16–19.
- [38] Marcus F. C. Alencar, Alberto B. Raposo, and Simone D. J. Barbosa. 2011. Composition of HCI evaluation methods for hybrid virtual environments. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, March 21, 2011. ACM, New York, NY, USA, 1237–1244. <https://doi.org/10.1145/1982185.1982457>
- [39] Sebastian Stadler, Henriette Cornet, and Fritz Frenkler. 2023. Assessing heuristic evaluation in immersive virtual reality—A case study on future guidance systems. *Multimodal Technologies and Interaction* 7, 2 (February 2023). <https://doi.org/10.3390/mti7020019>
- [40] Ivan David Tovar Siso, Kasper Amstrup Jensen, Stefan Norborg Eriksen, and Claus B. Madsen. 2022. Democratising mobilities design tools using VR - extended abstract. In *ACM International Conference Proceeding Series*, October 08, 2022. Association for Computing Machinery. <https://doi.org/10.1145/3547522.3547713>
- [41] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *Int J Aviat Psychol* 3, 3 (1993), 203–220. https://doi.org/10.1207/s15327108ijap0303_3
- [42] Presence Questionnaire. Retrieved February 12, 2024 from <https://marketinginvolvement.files.wordpress.com/2013/12/pq-presencequestionnaire.pdf>
- [43] Mores Prachyabrued and Christoph W. Borst. 2012. Virtual grasp release method and evaluation. *International Journal of Human-Computer Studies* 70, 11 (November 2012), 828–848. <https://doi.org/10.1016/j.ijhcs.2012.06.002>
- [44] Emmanuelle Chapoulie, Maud Marchal, Evanthia Dimara, Maria Roussou, Jean-Christophe Lombardo, and George Drettakis. 2014. Evaluation of direct manipulation using finger tracking for complex tasks in an immersive cube. *Virtual Real* 18, 3 (2014), 203–217. <https://doi.org/10.1007/s10055-014-0246-0>
- [45] Irene Katsouri, Aimilia Tzanavari, Kyriakos Herakleous, and Charalambos Poullis. 2015. Visualising and assessing hypotheses for marine archaeology in a VR CAVE environment. *Journal on Computing and Cultural Heritage* 8, 2 (March 2015). <https://doi.org/10.1145/2665072>
- [46] Dennis C. Frezzo and Muffie Wiebe Waterman. 2020. Immersive virtual reality for teaching problem solving in networking. In *Proceedings of 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2020*, December 08, 2020. Institute of Electrical and Electronics Engineers Inc., 968–971. <https://doi.org/10.1109/TALE48869.2020.9368450>
- [47] Andreas Junker, Carl Hutter, Daniel Reipur, Lasse Embø, Niels Christian Nilsson, and Rolf Nordahl. 2020. Virtual womb: Experiencing human sensory development from a fetal point-of-view in virtual reality. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, October 25, 2020. Association for Computing Machinery, 1–8. <https://doi.org/10.1145/3419249.3420159>
- [48] Diana Castilla, Azucena Garcia-Palacios, Juana Bretón-López, Ignacio Miralles, Rosa María Baños, Ernestina Etchemendy, Luis Farfallini, and Cristina Botella. 2013. Process of design and usability evaluation of a telepsychology web and virtual reality system for the elderly: Butler. *International Journal of Human Computer Studies* 71, 3 (2013), 350–362. <https://doi.org/10.1016/j.ijhcs.2012.10.017>
- [49] R. J. Hodes and D. A. Lindberg. 2002. Making your website senior friendly a checklist national institute on aging. National Institute on Aging and the National Library of Medicine (2002).
- [50] Jakob Nielsen. 2005. Ten usability heuristics. Retrieved February 28, 2023, from <http://designingwebinterfaces.com/6-tips-for-a-great-flex-ux-part-5>

- [51] Priska Breves. 2021. Biased by being there: The persuasive impact of spatial presence on cognitive processing. *Comput Human Behav* 119, (June 2021). <https://doi.org/10.1016/j.chb.2021.106723>
- [52] Mitsari Lucio-Alonso, Ricardo Mendoza-González, Huizilopoztli Luna-García, Zeoli A. Maldonado-Morales, Mario A. Rodríguez-Díaz and Francisco J. Luna-Rosas. 2019. A first version of design guidelines for virtual environments to support physical rehabilitation. In *ACM International Conference Proceeding Series*, September 11, 2019. Association for Computing Machinery, 157–161. <https://doi.org/10.1145/3364138.3364171>
- [53] Pedro Morillo, Juan M. Orduña, and José Duato. 2009. M-GRASP: A GRASP with memory for latency-aware partitioning methods in DVE systems. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans* 39, 6 (November 2009), 1214–1223. <https://doi.org/10.1109/TSMCA.2009.2025024>
- [54] Lewis Carter and Leigh Ellen Potter. 2016. Designing games for presence in consumer virtual reality. In *SIGMIS-CPR 2016 - Proceedings of the 2016 ACM SIGMIS Conference on Computers and People Research*, June 02, 2016. Association for Computing Machinery, Inc, 141–148. <https://doi.org/10.1145/2890602.2890626>
- [55] Eduardo Jose Goncalves De Oliveira, Francisco Paulo Magalhaes Simoes, and Walter Franklin M. Correia. 2017. Heuristics evaluation and improvements for low-cost virtual reality. In *Proceedings - 19th Symposium on Virtual and Augmented Reality, SVR 2017*, November 16, 2017. Institute of Electrical and Electronics Engineers Inc., 178–187. <https://doi.org/10.1109/SVR.2017.31>
- [56] Filipe Fernandes and Claudia Werner. 2019. Towards immersive learning in object-oriented paradigm: A preliminary study. In *Proceedings - 2019 21st Symposium on Virtual and Augmented Reality, SVR 2019*, October 01, 2019. Institute of Electrical and Electronics Engineers Inc., 59–68. <https://doi.org/10.1109/SVR.2019.00026>
- [57] Cristina Yuri Nakata Hara, Fernanda dos Santos Nogueira Goes, Rosangela Andrade Aukar Camargo, Luciana Mara Monti Fonseca, and Natalia Del Angelo Aredes. 2021. Design and evaluation of a 3D serious game for communication learning in nursing education. *Nurse Educ Today* 100, (May 2021). <https://doi.org/10.1016/j.nedt.2021.104846>
- [58] Benjamin S. Botha, Lizette de Wet, and Yvonne Botma. 2021. Experts' review of a virtual environment for virtual clinical simulation in South Africa. *Comput Animat Virtual Worlds* 32, 2 (March 2021). <https://doi.org/10.1002/cav.1983>
- [59] Shailey Minocha and Christopher Leslie Hardy. 2011. Designing navigation and wayfinding in 3D virtual learning spaces. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference*, November 28, 2011. ACM, New York, NY, USA, 211–220. <https://doi.org/10.1145/2071536.2071570>
- [60] Gareth W Young, Sam Stehle, Burcin Yazgi Walsh, and Ege Tiri. 2020. Exploring virtual reality in the higher education classroom: Using VR to build knowledge and understanding. *Journal of Universal Computer Science* 26, 8 (2020), 904–928.
- [61] Nathan Moore, Soojong Yoo, Philip Poronnik, Martin Brown, and Naseem Ahmadpour. 2020. Exploring user needs in the development of a virtual reality–Based advanced life support training platform: Exploratory usability study. *JMIR Serious Games* 8, 3 (July 2020). <https://doi.org/10.2196/20797>
- [62] Abdul Nasir Zulkifli, Mohd Fitri, Bin Yusoff, and Fadziana Faisal. 2014. Expert and user evaluations of persuasive hajj learning procedures courseware. *Knowledge Management International Conference (KMICE)* (August 2014), 12–15. Retrieved from <http://www.kmice.cms.net.my/>
- [63] Fred and Davis. 1987. User acceptance of information systems: the technology acceptance model (TAM). Michigan.
- [64] Paolo Belluco, Monica Bordegoni, and Samuele Polistina. 2010. Multimodal navigation for a haptic-based virtual assembly application. In *ASME 2010 World Conference on Innovative Virtual Reality*, January 01, 2010. ASME, 295–301. <https://doi.org/10.1115/WINVR2010-3743>
- [65] Etienne Van Wyk and M. R. Ruth De Villiers. 2014. Applying design-based research for developing virtual reality training in the South African mining industry. In *ACM International Conference Proceeding Series*, September 29, 2014. Association for Computing Machinery, 70–81. <https://doi.org/10.1145/2664591.2664627>
- [66] Ebba Thora Hvannberg, Gyda Halldórsdóttir, and Jan Rudinsky. 2012. Exploitation of heuristics for virtual environments. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction Making Sense Through Design - NordiCHI '12*, 2012. ACM Press, New York, New York, USA, 308. <https://doi.org/10.1145/2399016.2399065>
- [67] Roberto Munoz, Thiago Barcelos, and Virginia Chalegre. 2012. Defining and validating virtual worlds usability heuristics. In *Proceedings - International Conference of the Chilean Computer Science Society, SCCC*, 2012. 171–178. <https://doi.org/10.1109/SCCC.2011.23>
- [68] Muhamad Fajar, Yogi Udjaja, David, Andry Chowanda, Budi Juarto, and Yulianto. 2022. A comparative investigation of usability issues toward virtual reality implementation in a state-owned shipping service enterprise. *ICIC Express Letters, Part B: Applications* 13, 5 (2022), 545–552. <https://doi.org/10.24507/iciclb.13.05.545>
- [69] Nishant Bugalia, Subodh Kumar, Prem Kalra, and Shantanu Choudhary. 2016. Mixed reality based interaction system for digital heritage. In *Proceedings - VRCAI 2016: 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry*, December 03, 2016. Association for Computing Machinery, Inc, 31–37. <https://doi.org/10.1145/3013971.3014000>
- [70] Sahar Ebrahiminejad Motlagh Tehrani, Norziha Megat Mohd. Zainuddin, and Tannaz Takavar. 2014. Heuristic evaluation for virtual museum on smartphone. In *2014 3rd International Conference on User Science and Engineering (i-USER)*, September 2014. IEEE, 227–231. <https://doi.org/10.1109/IUSER.2014.7002707>

- [71] Andrea Ferracani, Daniele Pezzatini, Jacopo Bianchini, Gianmarco Biscini, and Alberto Del Bimbo. 2016. Locomotion by natural gestures for immersive virtual environments. In *AltMM 2016 - Proceedings of the 1st International Workshop on Multimedia Alternate Realities*, co-located with ACM Multimedia 2016, October 16, 2016. Association for Computing Machinery, Inc, 21–24. <https://doi.org/10.1145/2983298.2983307>
- [72] Janine Videva, Elena Marchiori, and Lorenzo Cantoni. 2019. Assessing usability and user experience of immersive web VR platforms for tourism destinations. *E-Review of Tourism Research* 17, 2 (December 2019).
- [73] Billy Sundström Langlet, Dorothy Odegi, Modjtaba Zandian, Jenny Nilstam, Per Södersten, and Cecilia Bergh. 2021. Virtual reality app for treating eating behavior in eating disorders: Development and usability study. *JMIR Serious Games* 9, 2 (April 2021). <https://doi.org/10.2196/24998>
- [74] Steven Vi, Tiago Silva, Frank Maurer, Frank Maurer User, and Tiago Silva Da Silva. 2019. User experience guidelines for designing HMD extended reality applications. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference*, September 02, 2019. Springer International Publishing, Paphos, Cyprus, 319–341. https://doi.org/10.1007/978-3-030-29390-1_18i
- [75] Abdulghafour Mohammad and Line Pedersen. 2022. Analysing the Use of Heuristics in a Virtual Reality Learning Context: A Literature Review. *Informatics* 9. <https://doi.org/10.3390/informatics9030051>
- [76] Chia-Chen Chen and Po-Hao Huang. 2023. The effects of STEAM-based mobile learning on learning achievement and cognitive load. *Interactive Learning Environments* 31, 1 (January 2023), 100–116. <https://doi.org/10.1080/10494820.2020.1761838>
- [77] Athanasios Christopoulos, Nikolaos Pellas, and Mikko Jussi Laakso. 2020. A learning analytics theoretical framework for stem education virtual reality applications. *Educ Sci (Basel)* 10, 11 (November 2020), 1–15. <https://doi.org/10.3390/educsci10110317>
- [78] Maged Soliman, Apostolos Pesyridis, Damon Dalaymani-Zad, Mohammed Gronfula, and Miltiadis Kourmpetis. 2021. The application of virtual reality in engineering education. *Applied Sciences* 11, 6 (March 2021), 2879. <https://doi.org/10.3390/app11062879>
- [79] Pavel Smutny, Marek Babiuch, and Petr Foltýnek. 2019. A review of the virtual reality applications in education and training. In *2019 20th International Carpathian Control Conference (ICCC)*, May 2019. IEEE, 1–4. <https://doi.org/10.1109/CarpathianCC.2019.876593>
- [80] Avellar, Gustavo Martins Nunes, and Barbosa, Ellen Francine. 2025. Heuristic evaluation of an affordable mobile VR application for K–12 programming education. In *Anais do Simpósio de Realidade Virtual e Aumentada (SVR)*. Sociedade Brasileira de Computação, Porto Alegre, RS, Brazil, 385–393. <https://doi.org/10.5753/svr.2025.40682>
- [81] Viol, Robson Santiago, Paula Cintra Fernandes, Iago Izidório Lacerda, Koda Gabriel Faria Melo, Arilton Junior de A. Nunes, André Schneider GM Pimenta, Cláudia Martins Carneiro, Mateus Coelho Silva, and Saul Delabrida. 2025. EduVR Collab: A tool for management immersive collaborative learning. *Journal on Interactive Systems* 16, 1, 120–136. <https://doi.org/10.5753/jis.2025.4276>
- [82] Dai, Chih-Pu, Fengfeng Ke, Nuodi Zhang, Alex Barrett, Luke West, Saptarshi Bhowmik, Sherry A. Southerland, and Xin Yuan. 2024. Designing conversational agents to support student teacher learning in virtual reality simulation: a case study." In *Extended abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–8. <https://doi.org/10.1145/3613905.3637145>
- [83] Baniqued, Paul Dominick E., Bremner, Paul, Johnson, Thomas, West, Andrew, Pulgarin, Erwin Jose Lopez, Flynn, David, Giuliani, Manuel, and Watson, Simon. 2024. Multimodal immersive digital twin platform for cyber–physical robot fleets in nuclear environments. *Journal of Field Robotics* 41, 5 (2024), 1521–1540. <https://doi.org/10.1002/rob.22329>
- [84] Chheang, Vuthea, Brian Thomas Weston, Robert William Cerda, Brian Au, Brian Giera, Peer-Timo Bremer, and Haichao Miao. 2024. A virtual environment for collaborative inspection in additive manufacturing. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–7. Association for Computing Machinery, New York, NY, USA, 1–7. <https://doi.org/10.1145/3613905.3650730>
- [85] Kontogiorgakis, Emmanouil, Emmanouil Zidianakis, Eirini Kontaki, Nikolaos Partarakis, Constantina Manoli, Stavroula Ntoa, and Constantine Stephanidis. 2024. Gamified VR storytelling for cultural tourism using 3D reconstructions, virtual humans, and 360 videos. *Technologies* 12,73. <https://doi.org/10.3390/technologies12060073>
- [86] Güner, Sevde, and Gül, Leman Figen. 2025. Evaluating virtual game design for cultural heritage interpretation: An exploratory study on arkeOyun. *Heritage* 8, 6. 208. <https://doi.org/10.3390/heritage8060208>
- [87] Almashmoum, Maryam, Payton, Antony, Johnstone, Emily, Cunningham, James, and Ainsworth, John. 2025. Understanding the views of health care professionals on the usability and utility of virtual reality multidisciplinary team meetings: Usability and utility study. *JMIR XR and Spatial Computing* 2 (2025), e60651. <https://doi.org/10.2196/60651>
- [88] Adjorlu, Ali, and Serafin, Stefania. 2024. Exploring the potentials of singing in immersive virtual reality to help children with social anxiety. In *Proceedings of the Sound and Music Computing Conference (SMC)*. Sound and Music Computing Network, 158–164. <https://doi.org/10.5281/zenodo.14337091>
- [89] Othman, Mohd Kamal, Mat, Rahimah, and Zulkiply, Norehan. 2025. WVREA heuristics: A comprehensive framework for evaluating usability in wearable virtual reality educational applications (WVREA). *Education and Information Technologies* 30, 10, 13599–13662. <https://doi.org/10.1007/s10639-024-13234-5>
- [90] Rauschnabel, Philipp A., Reto Felix, Chris Hinsch, Hamza Shahab, and Florian Alt. 2022. What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior* 133 (2022), Article 107289. <https://doi.org/10.1016/j.chb.2022.107289>