

A Systematic Review on Extended Reality-Mediated Multi-User Social Engagement

Yimin Wang ^{1,2}, Daojun Gong ^{1,3}, Ruowei Xiao ^{1,*} , Xinyi Wu ¹ and Hengbin Zhang ¹

¹ Laboratory of METAPHOR, School of Design, Southern University of Science and Technology, Shenzhen 518055, China; wangym2022@mail.sustech.edu.cn (Y.W.); iamdjgong@mail.nwpu.edu.cn (D.G.); 12333223@mail.sustech.edu.cn (X.W.); 12333217@mail.sustech.edu.cn (H.Z.)

² Faculty of Humanities and Arts, Macau University of Science and Technology, Macau 999078, China

³ Institute of Flexible Electronics, Northwestern Polytechnical University, Xi'an 710060, China

* Correspondence: xiaorw@sustech.edu.cn

Abstract: The metaverse represents a post-reality universe that seamlessly merges physical reality with digital virtuality. It provides a continuous and immersive social networking environment, enabling multi-user engagement and interaction through Extended Reality (XR) technologies, which include Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). As a novel solution distinct from traditional methods such as mobile-based applications, the technical affordance of XR technologies in shaping multi-user social experiences remains a complex, multifaceted, and multivariate issue that has not yet been thoroughly explored. Additionally, there is a notable absence of mature frameworks and guidelines for designing and developing these multi-user socio-technical systems. Enhancing multi-user social engagement through these technologies remains a significant research challenge. This systematic review aims to address this gap by establishing an analytical framework guided by the PRISMA protocol. It analyzes 88 studies from various disciplines, including computer science, social science, psychology, and the arts, to define the mechanisms and effectiveness of XR technologies in multi-user social engagement. Quantitative methods such as descriptive statistics, correlation statistics, and text mining are used to examine the manifestation of mechanisms, potential system factors, and their effectiveness. Meanwhile, qualitative case studies identify specific measures by which system factors enhance multi-user social engagement. The study provides a pioneering framework for theoretical research and offers practical insights for developing cross-spatiotemporal co-present activities in the metaverse. It also promotes critical reflection on the evolving relationship between humans and this emerging digital universe.



Citation: Wang, Y.; Gong, D.; Xiao, R.; Wu, X.; Zhang, H. A Systematic Review on Extended Reality-Mediated Multi-User Social Engagement. *Systems* **2024**, *12*, 396. <https://doi.org/10.3390/systems12100396>

Academic Editor: Mariusz Salwin

Received: 8 August 2024

Revised: 19 September 2024

Accepted: 23 September 2024

Published: 26 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The concept of the metaverse, initially introduced in the sci-fi novel *Snow Crash* [1], has evolved into a dynamic virtual environment that complements the real world. It drives socio-cultural and economic advancements across various sectors, such as family entertainment, telemedicine, and non-fungible token markets [2–4]. Significant developments, such as Facebook's rebranding to Meta in October 2021 [5], have fueled public interest. Projections estimate that the market will reach a valuation of USD 507.8 billion by 2030, with an annual growth rate of 37.73% starting from 2024 [6]. However, the metaverse also faces challenges, including harassment, avatar sexualization, data exploitation, and unregulated gambling [7]. These issues, both positive and negative, are closely linked to multi-user social engagement (MSE). Extended Reality (XR), as the supportive platform for MSE, encompasses Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), offering a more immersive MSE solution. XR allows users to engage in

hyper-spatiotemporal communication and multi-sensory self-expression through technologies such as head-mounted displays (HMD), tangible user interfaces (TUI), and tracking systems. These technologies help users achieve higher social presence, motivation, and enjoyment in MSE [2]. These modes far exceed conventional screen-based interactions, which potentially lead to inadequate self-awareness, low interactivity, and unnuanced emotional expression [2,7,8].

Recent research on XR-mediated MSE spans multiple disciplines, including social sciences, computer science, and psychology, extending into domains such as rehabilitation, business, and education. Despite these advancements, significant research gaps remain. The field lacks universally recognized frameworks and guidelines, complicating the articulation of technology-mediated mechanisms and impeding practical application. The technical affordances of XR and their impact on MSE engagement are still complex and underexplored. Understanding which system factors (SFs) of XR enhance or diminish activities is crucial. Additionally, fragmented cases do not adequately demonstrate the effectiveness of XR-mediated MSE, affecting the sustainability and replicability of solutions. To address these gaps, the following research questions (RQs) are proposed.

RQ1. What is the mechanism of XR for mediating MSE?

- RQ1.1. What are the cognitive–behavioral mechanisms of MSE and their manifestations?
- RQ1.2. What are the technology-mediated mechanisms of XR and their manifestations?
- RQ1.3. What are the strongly correlated SFs that influence MSE?

RQ2. How has the effectiveness of XR-mediated MSE been proven?

- RQ2.1. What evaluation settings have been established for XR-mediated MSE?
- RQ2.2. What is the performance of these evaluations?

By answering these RQs, this study offers substantial implications for the XR-mediated MSE field:

- Theoretical implications. By addressing RQ1.1, RQ1.2, and RQ1.3, this study proposes a novel analytical framework for XR-mediated MSE. This framework integrates comprehensive elements supported by interdisciplinary theories and is validated by sample data, providing an insightful foundation for future development. The framework highlights the critical role of engagement as a prerequisite for social processes and outcomes. It urges researchers to revisit and update their theoretical perspectives to thoroughly examine future empirical studies on XR-mediated MSE;
- Practical implications. By answering RQ2.1 and RQ2.2, this study analyzes validation methods and the current performance of XR-mediated MSE. A case study is also provided in Section 5.2.3 Significant Mediated Relationships, further explaining strongly correlated SFs and their mediation mechanisms to MSE. These insights emphasize both the potential advantages and limitations of current empirical research in the field. The findings offer a comprehensive blueprint for effective design and implementation for practitioners in both industry and academia, including design researchers, metaverse designers, and developers. For example, the results concerning SF-mediated mechanisms can guide the metaverse product pipeline by supporting knowledge graphs of cross-spatiotemporal or co-present activities;
- Societal implications. Although the metaverse is not yet fully ubiquitous, this study explores the significance of XR as an emerging socio-technical system for future human social activities through comprehensive answers to RQ1 and RQ2. The findings are broadly applicable, potentially attracting interdisciplinary and cross-sectoral interest. This study encourages further reflection on the evolving relationship between people and the metaverse, supporting innovative perspectives for transforming traditional industries (e.g., tourism to cloud tourism).

2. Concepts and Analysis Framework

2.1. Cognitive–Behavioral Mechanism of Multi-User Social Engagement (MSE)

Social interaction (SI) stems from human nature as interconnected beings. It involves developing psychosocial bonds with others, facilitated by co-coupling that creates a self-sustaining network of symbol exchange—not inherent information flows [9–12]. Social habitualization cultivates human cultural attributes, rules of thumb, and resource exchanges that shape human worldview [13,14]. Engagement, a multidimensional concept, involves attributes such as connection, interaction, participation, and involvement aimed at achieving outcomes at individual, organizational, or social levels [15]. Social-level engagement is characterized by cognitive forms (e.g., shared knowledge) and behavioral forms (e.g., collective action, group participation) within a socially situated system [15].

Xiao et al.’s 3-axis *User Engagement Construct* delineates social engagement across varying scales [16]. (1) One-on-one engagement, involves direct exchanges between two individuals, driven by principles of *Social Exchange* [17], *Attachment* [18], and *Interaction Ritual Chains* [19], emphasizing private expression and rich meaning [15]. (2) Group-level engagement, ranging from dyadic to larger public groups, is shaped by specific roles within situations such as healthcare, education, and family. According to Luhmann’s *Social Systems Theory* [20], functional differentiation arises from selective pressure under temporalized complexity. This self-referential nature enables groups to self-regulate independently of external influences. Scalability relies on having multiple participants to form a functioning group. Disruptions can impact the entire group dynamic, as suggested by *Social Functionalism Theory* [21] and *Social Capital Theory* [22]. Another characteristic is the deeper penetration of social resources, as ‘power and control underpin all organizational–social relationships’ [15]. (3) Public engagement, represents a macro-level phenomenon characterized by loose connections, minimal mediation, and the direction of information flow [23]. This scalability targets an unspecified audience and is crucial for socialization [24]. For instance, wisdom-crowdsourced activities that exhibit higher levels of relational equality can enhance societal justice, decision-making quality, and well-being [25,26]. All in all, different scales of MSE illustrate a process in which multiple participants co-construct a network-based reality through shared cognition, behaviors, and interconnected experiences.

Understanding the cognitive–behavioral mechanisms of MSE is essential for grasping its antecedents, processes, and outcomes, thereby clarifying the mediating relationship between XR and MSE. Since Bales and Strodtbeck’s foundational work in 1951 [27], the field has evolved significantly. It has moved from situation-explained models that focus on aspects such as motivation, goal orientation, problem-solving strategies, social episodes, settings, and *face-to-face* (*FtF*) social patterns [28–35]. The field now includes algorithm-driven approaches, such as *Relational Event Modeling* [36]. However, these models often fall short of accurately analyzing MSE as they emphasize process details and specific social situations (e.g., Krause’s taxonomy for joint-working [28]) rather than the concepts of ‘engagement’ and ‘multi-user’ itself. Consequently, most frameworks may be too specific and limited in their application to capture all characteristics of SI. Concepts from the ‘3-axis *User Engagement Construct*’ [16] were incorporated. For example, motivational engagement is integrated into the actor’s attributes and establishes labels for autonomy. Additionally, common analytical components in multi-user situations, such as topology, clustering, and familiarity, have also been incorporated.

A major concern in constructing a framework for MSE is having high-level primitive descriptions that encompass various social situations [37–39]. To address this gap, a new framework grounded in MSE phenomena is proposed. Drawing on the *APRACE Model* [39], which breaks down SI into six dimensions (actor, partner, relation, activities, context, and evaluation). These dimensions were streamlined into three categories: actors (combining actor and partner), relations, and contexts (combining activities and context), with evaluation integrated directly into the indicators. This refinement enhances the framework’s objectivity and minimizes situation-specific biases. Detailed explanations of the framework’s components are discussed in Sections 2.1.1–2.1.3.

2.1.1. Actors: Autonomy, Affect, and Cues

'Actors', as defined in *Actor-network Theory (ANT)* [40] and *Media Equation Theory* [41], are any beings with ascribed actions, framing human interactions within a *Social Ontology* [42]. This term aptly captures the proactive engagement of MSE participants. *Symbolic Interactionism* asserts that group dynamics must be understood in terms of actors' action attributes—components that activate individuals' meaning construction [43]. These indispensable attributes are as follows:

- (1) Autonomy. Rooted in *Self-Determination Theory (SDT)* [44], autonomy in MSE arises from the interplay of personal agency and external influences [45]. SDT characterizes motivation as a spectrum, ranging from motivation through extrinsic motivation to intrinsic motivation [46]. This spectrum represents various motivational types for MSE—from pure intrinsic motivation, driven by personal enjoyment, to a blend of extrinsic and intrinsic motivations, where external rewards enhance personal interest. Notably, solely extrinsic motivation is uncommon in MSE contexts;
- (2) Affect. Affect encompasses a wide array of emotions sustained over time beyond transient reactions. Myers categorizes affect into three dimensions: physiological arousal, expressive behaviors, and conscious experience, with arousal as the core element [47]. While tools such as the *Self-Assessment Manikin (SAM)* represent arousal levels [48], they may not capture the full spectrum of affective states in MSE. These states can be intensive, sporadic, or a mix of both;
- (3) Cues. Essential for MSE, cues transmit information, intentions, and emotions, traditionally classified as verbal or non-verbal [49,50]. Verbal cues are essential in the interaction between human users and humanoid interfaces. Non-verbal cues are equally important. The importance of non-verbal cues becomes especially evident in situations where non-verbal engagement behaviors conflict with verbal behaviors [51]. This conflict can lead to difficulties in maintaining communication. However, this binary classification may lack the granularity for using XR systems. Incorporating insights from studies on *Embodied Agents* [52], *Task Coordination and Communication* [53], and *Conversational Agents* [54], an expanded classification of cues has been developed. This classification includes verbal communication, facial expressions, gestures, touch, posture, object control, and material sharing.

2.1.2. Relations: Topology, Familiarity, Clustering, and Coopetition

'Relations' are fundamental to MSE as they shape the actors' sense of 'self' through mutual communication. Social cognition extends beyond individual cognitive processes. It involves interactions that do more than just provide a context; these interactions can enhance or even substitute for individual cognitive functions [12]. This aligns with the *Interactionist perspective* [43] and is supported by *Structural Functionalism*, which underscores the influence of relation structures on societal functions [55,56]. The concept of 'relations' in MSE encompasses four key parts:

- (1) Topology. Analyzed using *Graph Theory*, topology describes the spatial structure of social relations, presenting society as an undirected network to reveal various social configurations [57]. Understanding the representational significance of topology in social relationships is crucial. For example, in a tree topology, a guidance agent acts as scaffolding. It provides instructions and emotional support to help learners complete tasks they cannot manage independently. This process establishes a tree-like relationship between the agent and the learners [58]. This study synthesizes *Local Area Network (LAN) Topologies* [59] to outline basic types: Point to Point: One-on-one interactions, e.g., private conversations. Star: Central communication hub, e.g., lectures and concerts. Mesh: Complex interconnections, e.g., research collaborations. Bus: Linear route (actors' cognitive-behavioral focus) with their immediate entry/exit, e.g., exhibitions. Tree: Hierarchical structures, e.g., leadership or teacher-student relationships. Highly Hybrid: Combination of three or more topologies;

- (2) Familiarity. This represents the depth of relationships, or social connectivity, similar to the core measurement content of the *Personal Acquaintance Measure (PAM)* [60]. Familiarity captures nuances of relationship duration, frequency, and openness to disclosing personal information [61]. Categories include Familiar-oriented: Stable connections; Unfamiliar-oriented: No prior relationship; and Mixed: Combination of both;
- (3) Clustering. Drawing from the *Two-step Flow of Communication Theory*, opinion leaders, serving as key nodes, facilitate greater clustering of efficient information and promote two-step communication, especially in large events [62]. This influences social opinions and behaviors and enhances network vitality. Clustering describes group formations with dense internal and sparse external ties. Origins of clustering include Spontaneous: Shared interests; Organized: Organizational affiliations. Personal influential: Influence from opinion leaders; Combination: Mix of types; All: Inclusive of all types; and None: Subtle manifestation;
- (4) Coopetition. According to *Game Theory* [63] and *Social Interdependence Theory* [64], actors exhibit both self-interested and altruistic behaviors, characterized as coopetition—a fusion of cooperation and competition [63,65]. This encompasses cognitive processes (e.g., team cognition), motivational processes (e.g., cohesion), and behavioral processes (e.g., coordination), contributing to performance and effectiveness [66]. Forms include Competition, Collaboration, Collaborative Competition, Competitive Collaboration, Neutral, and Combination [67].

2.1.3. Contexts: Spatial and Temporal Conditions

Giddens's *Constitution of Society* emphasizes that personal agency is fundamentally shaped by intentionality within social 'contexts' [68,69]. Proponents of *Situated Theories of Learning* and *Distributed Cognition* assume the inseparability of social context from individual agentic action [70–72]. They focus on the reproduction of cultural practices through situated SI and participation processes. Each social and organizational level possesses distinct forms of space and time, shaping social practices by defining their spatial and temporal characteristics [73]. Understanding these characteristics is crucial for grasping the strength of connections between different social practices [74].

- (1) Spatial conditions. Abbott proposes that this is a specifically Chicago School insight: 'One cannot understand social life without understanding the arrangements of particular social actors in particular social times and places... Social facts are located' [75,76]. The sub-classification of spatial conditions [8], based on MSE situations, effectively represents these contextual attributes. Spatial conditions are categorized into Collocated, Remote, and Mixed. Collocated settings involve direct, co-present communication, allowing for immediate feedback and richer non-verbal communication. Remote settings involve physical separation and rely on technology, which introduces different dynamics;
- (2) Temporal conditions. Social time is multi-dimensional, irregular, and multidirectional, with social time following a rhythm specified by human activity [77]. Emirbayer and Mische define agency as '*temporally constructed engagement by actors*' [72]. Temporal conditions are classified as Synchronized (real-time interaction), Asynchronized (delayed interaction), and Hybrid (a mix of both). Analyzing these temporal conditions helps us understand the dynamics and effectiveness of interactions [77].

2.2. Extended Reality (XR) Systems

The 'X' in XR stands for a range of emerging reality formats, indicating an open-ended description of future spatial computing technologies, also known as *xReality* [78]. Over the past quarter-century, Milgram and Kishino's *Reality-Virtuality Continuum* [79] has become a foundational work in the field, cited thousands of times. Most researchers in the XR community still acknowledge the utility of Milgram and Kishino's continuum for classifying systems [80]. Until an alternative framework gains unanimous acceptance, their

model continues to be valuable for knowledge categorization and dissemination. Therefore, this study utilizes the *Reality-Virtuality Continuum* [79] to categorize XR interfaces by the degree of computer-generated virtual objects integrated into the physical world, including (see Figure 1) (a) Augmented Reality (AR), which enhances real-world environments with digital elements, allowing interactive virtual overlays while retaining physical context awareness; (b) Mixed Reality (MR), which integrates real and virtual worlds, where physical and digital entities coexist and interact in real time; and (c) Virtual Reality (VR), which provides a fully immersive virtual environment, isolating the user from the physical world. Further definitions of the categories can also be found in Figure 1, as explained by Tremosa and the *Interaction Design Foundation* [81]. It is worth noting that the emergence of multi-interface XR systems is observed, although there is currently no classification model for them. To address this, a classification for these systems has been designed. An XR system might include a single interface type, termed a single-interface system (e.g., solely AR). Alternatively, it might combine several interfaces, referred to as a multi-interface system (e.g., AR combined with VR).

● The Definition of XR Systems

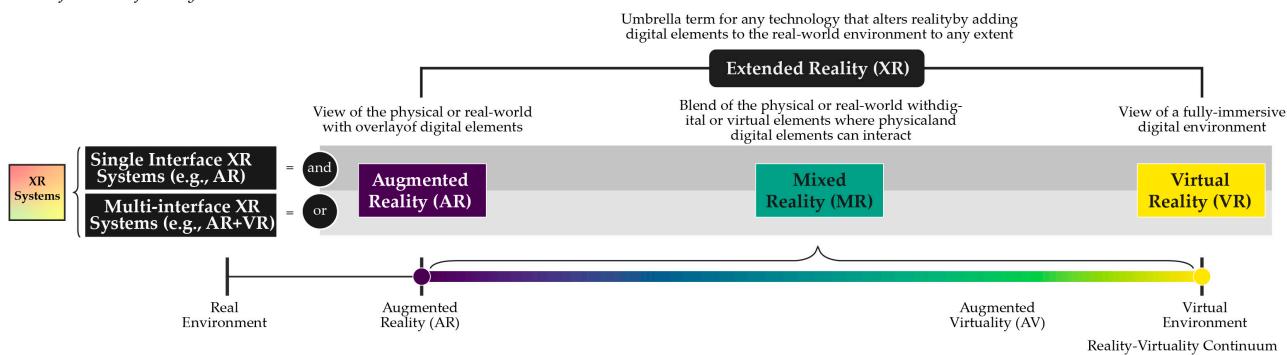


Figure 1. The Definition of XR Systems. This study uses the classification method and concepts proposed by the Reality-Virtuality Continuum [79] and Tremosa and the Interaction Design Foundation [81] in defining the XR systems.

In revisiting related models that describe human–computer interaction (HCI) elements, several were examined. These include the *PACT Model* [82], which evaluates solutions through ‘people’, ‘activities’, ‘contexts’, and ‘technologies’; the *SMPC Model* [83], which considers ‘subjects’, ‘modes’, ‘purposes’, and ‘contexts’; and Silver’s model [84], a variant of Moggridge’s *Designing Interactions Model* [85], which includes five design languages (‘words’, ‘visual representations’, ‘physical objects or space’, ‘time’, and ‘behavior’). Several potential SFs from these dimensions were identified. The *SMPC Model* [83], in particular, provides a comprehensive framework for understanding interaction as mutual or reciprocal actions between entities with specific goals within certain contexts. This model helps define interaction in terms of ‘subjects’ (e.g., who interacts with us? AI or human?), ‘modes’ (e.g., how do these subjects interact? Using avatar or not?), ‘purposes’ (e.g., why is interaction taking place? Utilizing task-based absorption to enhance group dynamic in MSE?), and ‘contexts’ (e.g., where is interaction happening? How is users’ activity range?). For example, the ‘subjects’ dimension is crucial because it includes both humans and AI. This emphasizes the need to consider AI as an interaction target. Unlike humans, AI does not share personal values or experiences, which affects human-mode MSE interactions [86–88]. Thus, a single SF should summarize these categories and be named ‘interaction targets’. The establishment of other SFs follows this pattern. The construction process for each SF is detailed in Sections 2.2.1–2.2.6.

2.2.1. Interaction Targets

Drawing from the *Computers Are Social Actors* (CASA) [89], the authenticity of interaction targets shapes social perceptions and influences behavioral realism [90,91]. Previous research highlights differences in these aspects between AI agents and humans (e.g., behavioral realism and social appropriateness) [86–88]. Consequently, interaction targets are categorized into (a) Human: solely actual human participants; (b) Human+AI: A mix of humans and AI agents; a separate tag for pure AI interactions (e.g., human–AI dyads without real human presence) was not established. This is because the study emphasizes multi-user interactions, focusing specifically on the value of human–human interactions. Pure AI as an interaction target is beyond the scope of this study.

2.2.2. Avatars

Avatars serve as embodied digital identities, allowing users to engage with the environment and virtual objects from multiple perspectives, such as the third-person view [2]. Avatars enhance visual realism by replicating social cues from real humans [88]. Their effectiveness in improving psychological aspects such as attention, performance, presence, trust, involvement, satisfaction, and self-disclosure makes avatars an indispensable SF [90]. Analyzing avatar settings involves several aspects: (a) Usage: Whether avatars are used in the XR system indicates if interactions are through real, physical bodies or avatars. (b) Visual Form: Ranges from partial (e.g., a hand, upper body) to full body representation. (c) Symmetry: The symmetry in avatar usage between oneself and partners, such as having no avatar for oneself while the other has an avatar with partial visual representation.

2.2.3. Activity Range

Interaction in XR environments does not require users to be stationary. Physical movement is transferred into XR environments through positional and rotational tracking. This allows users to activate their entire bodies [2]. Activity range measures involvement with it, defined through the affordable extent of subjective freedom for physical activity [92], and can be quantified using the following modes: (a) Stationary: Limited to head and arm movement without positional changes; (b) Table-sized. Allowing for more extensive interaction within a confined area; (c) Room-sized. Permitting free movement and interaction within a physical space that matches the dimensions of a typical room, providing users with enough area to perform various actions safely; (d) Pervasive. Removing physical constraints.

2.2.4. Absorption

Absorption in users is driven by immersion, which refers to the degree to which a user feels cognitively and behaviorally transported to an alternative, synthetic world [2,93]. It can be assessed across multiple dimensions: (a) Narrative. Deep empowerment in story-based events; (b) Strategic. Focused on planning and strategy; (c) Tactical. It involves physical and sensorimotor tasks that enhance immersion through sensorimotor and physiological responses. Notably, this study employed a broader definition of tactical immersion, distinguishing it more clearly from strategic immersion. Because XR systems emphasize embodiment more than traditional CMC devices, analyzing the tactical dimension can reveal specific characteristics of XR [2]. Hence, we extend the definition by Adams and Rollings. They often link strategic and tactical immersion together, describing it as “*a state of intense preoccupation with observation, calculation, and planning or with swift responses to obstacles*” [94]. We expand this definition to encompass a broader range of tactical immersion. This includes immersion arising from bodily movement and physiological responses, beyond the narrow, immediate immersive experience typically associated with hand-eye coordination. Additionally, it is noteworthy that all these immersion classifications extend Ryan’s concept of ‘*Ludic Immersion*’. In other words, the antecedents of these immersions are considered to be generated by researchers based on goal-oriented tasks, defined as ‘*a state of intense absorption in the task currently being performed*’ [95]. Specifically, immersion in VR is largely determined by the quality of the virtual environment. This includes factors

such as modality, visual representation, interactivity, haptic feedback, audio quality, depth cues, video, and display [96]. Task flow also plays a significant role. In contrast, AR and MR immersions are not primarily based on the virtual environment. Instead, they depend more on task flow. Utilizing task-stimulated immersion allows for a more standardized analysis of absorption.

2.2.5. Input and Output Devices

Unlike conventional CMC devices, XR interfaces are more focused on activating the users' motor and perceptual systems. This focus facilitates a more nuanced exchange of information elements, such as expression, memory, emotion, and language [97]. Input devices are tools for users to submit data and commands, such as hand controllers, motion trackers, and voice recognition systems, translating human intentions into actions within the system [98]. Output devices such as HMD, projectors, and wearable haptics are vital for delivering system responses and ensuring user perception.

2.2.6. Sensory Feedback

XR can deliver computer-generated stimuli to all the exteroceptive senses, that is, those senses that respond to stimuli from outside the body [80]. By incorporating sensory feedback beyond the usual visual and auditory aspects, users' immersion, presence, agency, and control are deepened, transforming interactions to closely mimic real-world experiences [99,100]. Key categories of sensory feedback include [101] (a) Visual: Dynamic, immersive graphics that respond to user movements, intensifying the realism of the virtual world. (b) Auditory: Soundscapes that reflect ambiance, aiding in accurate intent, understanding, and expression. (c) Tactile: Varied textures and haptic feedback that simulate touch, enhancing memory sensation and realism. (d) Olfactory: Scents that trigger vivid recollections of past experiences, deepening emotional and contextual engagement. (e) Gustatory: Engages the sense of taste, providing a deeper embodied perception of information.

2.3. Overview of the XR-Mediated MSE Analysis Framework

Contemporary *Mediation Theory* posits that technologies have their own form of 'agency' for human activities and are mutually constitutive with human subjects [44,102]. Ihde's approach to *Mediation Theory* includes the intentionality, embodiment, and perception of human activities. From this perspective, the focus should not only be on the technology itself (RQ1), but also on how mediation influences perceptions of events and the technology itself, particularly its effectiveness (RQ2) [103]. Sections 2.1 and 2.2 construct the cognitive-behavioral mechanism of MSE (Figure 2A) and the technology-mediated mechanism of XR (Figure 2B) by reviewing multidisciplinary theories and frameworks. These mechanisms incorporate various components (e.g., autonomy, affect, cues, topology, familiarity, clustering, competition, spatial conditions, temporal conditions) and SFs (e.g., interaction targets, avatars, absorption, activity range, input devices, output devices, sensory feedback). Discrete labels were designed to quantify these variables (e.g., (a-)synchronized and mixed communication for temporal conditions) in the analysis framework (Figure 2). It is hypothesized that the properties of components (e.g., discrete labels and their quantitative distributions) characterize their cognitive-behavioral and technology-mediated mechanisms, addressing RQ1.1 and RQ1.2. Examining correlations between these components and SFs addresses RQ1.3, i.e., extracting strongly correlated SFs. Additionally, this standardized classification provides insights into validating the effectiveness analysis in RQ2.

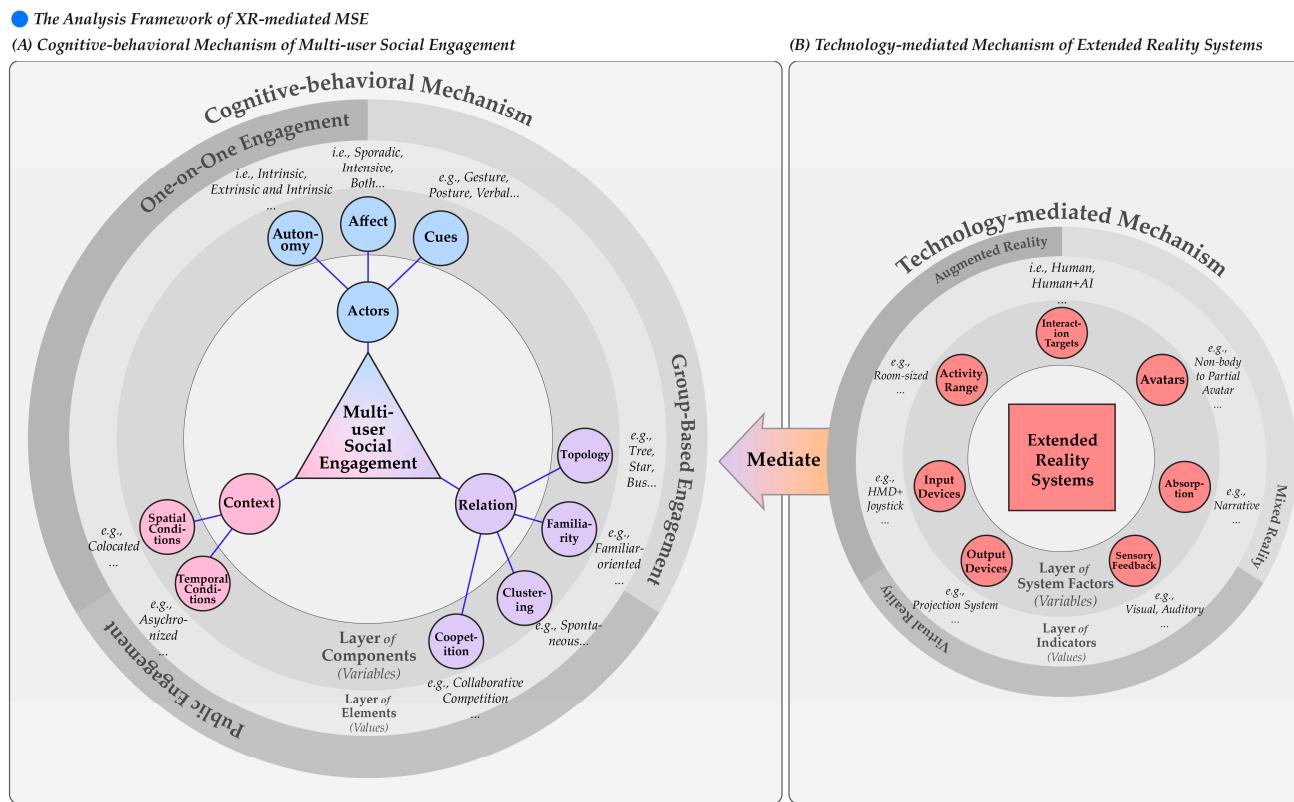


Figure 2. The Analysis Framework of XR-mediated MSE. (A) Cognitive–behavioral Mechanism of Multi-user Social Engagement. (B) Technology-mediated Mechanism of Extended Reality Systems.

3. Methodology

Research samples were collected following the *Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) Protocol* [104]. A comprehensive database, Scopus [105], was mainly used for searching since it also contained results indexed in other specific databases such as the *Web of Science*. The results were solidified to be replicable, rigorous, and transparent by setting standard searching algorithms. Based on inputting topic-related searching strings, other parameters were configured to encompass outcomes confined within the publication type of conference and journal papers. Book chapters were not collected because they tend to provide overviews and summaries. These sources may lack detailed experimental design and data analysis specifics, which may not be conducive to comprehensive quantitative analysis. Furthermore, samples were refined to the disciplines of computer science, social science, psychology, and arts due to the consideration of domain and thematic relevance. The search query string was as follows:

(TITLE-ABS-KEY (“extended reality”) OR TITLE-ABS-KEY (“augmented reality”) OR TITLE-ABS-KEY (“mixed reality”) OR TITLE-ABS-KEY (“virtual reality”) OR TITLE-ABS-KEY (“immersive”) OR TITLE-ABS-KEY (“virtual environment”) AND TITLE-ABS-KEY (“public”) OR TITLE-ABS-KEY (“multi-user”) OR TITLE-ABS-KEY (“multiple”) OR TITLE-ABS-KEY (“multiperson”) OR TITLE-ABS-KEY (“commu*”) OR TITLE-ABS-KEY (“social interaction”) AND TITLE-ABS-KEY (“engagement”) AND (LIMIT-TO (SRCTYPE, “j”) OR LIMIT-TO (SRCTYPE, “p”)) AND (LIMIT-TO (DOCTYPE, “cp”) OR LIMIT-TO (DOCTYPE, “ar”)) AND (LIMIT-TO (SUBJAREA, “COMP”) OR LIMIT-TO (SUBJAREA, “SOCI”) OR LIMIT-TO (SUBJAREA, “PSYC”) OR LIMIT-TO (SUBJAREA, “ARTS”)) AND (LIMIT-TO (LANGUAGE, “English”))

By June 2024, a total of 1399 search results were accumulated. Details of the flow chart and screening process can be found in Supplementary Files S1 and S2. During the initial screening based on titles and abstracts, 1015 results were excluded due to topic irrelevance. Specifically, samples that were not XR-based interfaces or did not involve multi-user social

patterns, such as interactions with AI systems alone, were ruled out. This left us with 384 results for detailed content screening. In the second screening round, 80 papers were identified as relevant to this study. A total of 271 papers were excluded due to the lack of identifiable descriptions of research content, methods, or experimental results. Additionally, 23 papers were excluded for being review-related contributions, such as case studies. Six papers were excluded for containing repetitive prototypes with no evident iteration. Four papers were excluded due to the lack of full-text access. Through snowballing, 8 additional papers were integrated, resulting in a final sample pool of 88 papers.

These 88 papers [23,46,51,106–190] were then coded according to the analysis framework outlined in Section 2.3 (see Figure 2). To enhance the reliability of the coding process, all authors conducted the coding independently. Any discrepancies were resolved through discussions until a 100% consensus was reached (see Supplementary File S2). A detailed examination of the statistical distribution of each metric was conducted based on RQs, such as identifying strongly correlated SFs for mediating MSE (to answer RQ1.3).

4. Results

4.1. Bibliometric Information

4.1.1. Publication Trends and Topics

Our analysis of the publication years and types of the accepted samples revealed a significant trend, with a notable increase in publications post-2017 (76/88, 86.4%). Data from 2024 may be incomplete due to indexing delays at the time of search. The earliest publications date back to 2002. A total of 23.9% of the works (21/88) spanned multiple disciplines. Computer science accounted for the predominant portion at 60.2% (53/88), followed by social sciences (27/88, 30.7%), arts (8/88, 9.1%), and psychology (5/88, 5.7%). This reflects the field's technological emphasis and cross-disciplinary collaboration.

Analysis of topics was conducted after the data pre-processing, involving deduplication, merging, alignment, and splitting of synonyms of the samples' keywords. Overall, 49 keywords (N) were identified as common concerns, each with a frequency (f) greater than 2. Topics were categorized into two groups: MSE-related and XR-related.

Within the MSE-related topics (Figure 3), the most notable focus was on target groups. Keywords in this category included 'Residents' (f = 17), 'Visitors' (f = 6), 'Older adults' (f = 7), 'Children' (f = 9), 'Patients' (f = 11), and 'Students' (f = 2). Samples emphasize the engagement of groups with civic identity such as residents and visitors. Additionally, it delves into minor groups, particularly aging populations [46,164], and specific patient groups, including those with autism [156], aphasia [129], dementia [127,136,184], stroke [149,158], and other unspecified conditions [164,179,185]. This underscores the current topic's emphasis on embracing technological equity and inclusiveness for all groups in MSE. Collective behaviors were another significant focus, with key terms such as 'Cooperation', 'Co-Creation', and 'Open Innovation'. Quality of socialization was also prominent, encompassing general terms such as 'Engagement' (f = 11) and more specific ones such as 'Embodied Cognition Theory (ECT)' (f = 5), 'Motivation' (f = 3), and 'Psychophysiology' (f = 3). Particularly noteworthy is the term 'Asymmetric Engagement' (f = 3); it corresponds to studies that address the challenges of isolated participation caused by limited XR equipment [114,167,190]. This indicates the significance of cross-spatiotemporal MSE, enabling worldwide engagement on equal footing, unbound by geographical restrictions—an increasing focus.

● MSE-related Topics of Common Concern

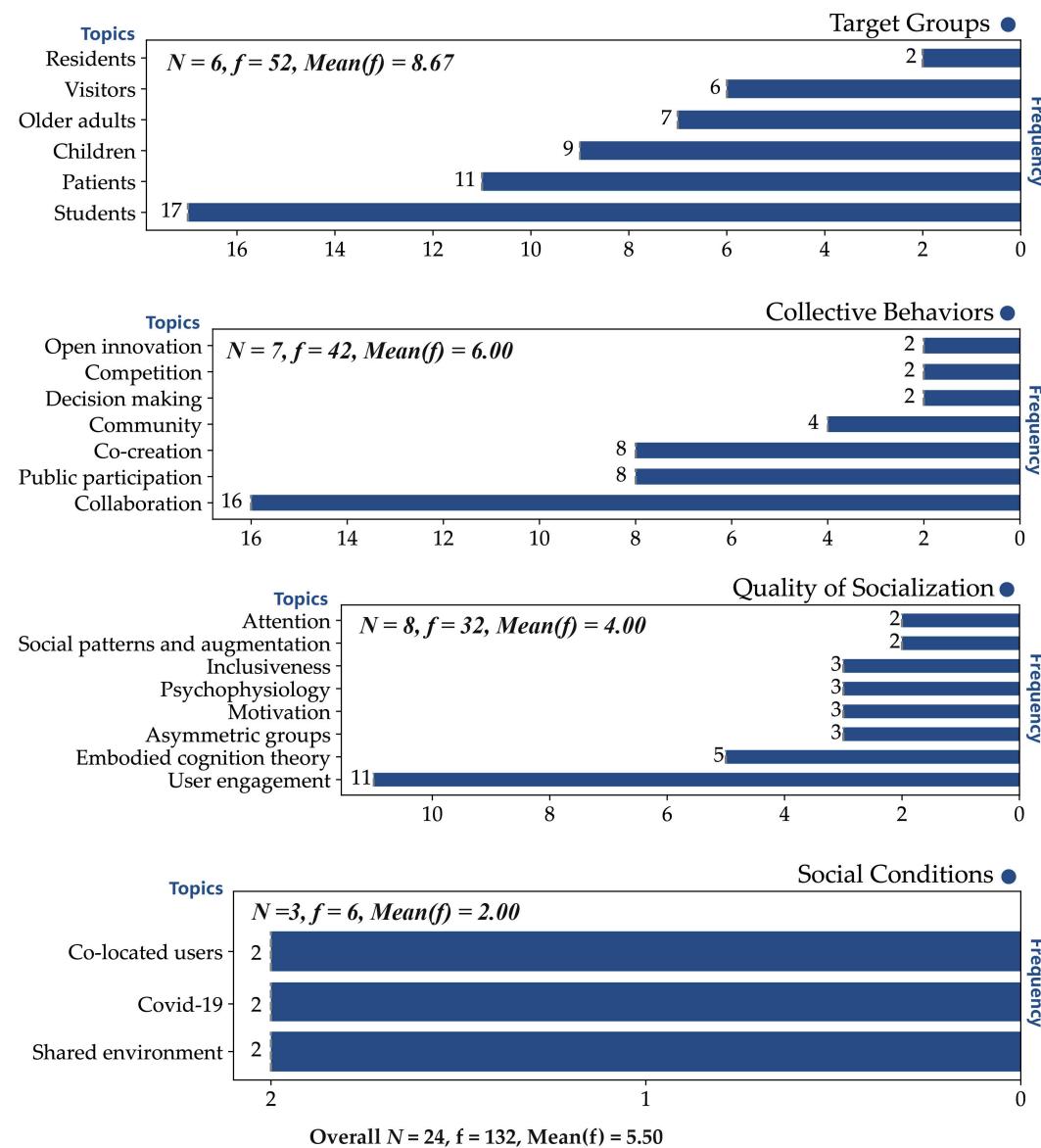


Figure 3. MSE-related Topics in the Collected Samples.

In the XR-related topics (Figure 4), the most significant feature was multi-technical integration, represented by keywords such as '*Tangible User Interface*' ($f = 6$), '*Web Services*' ($f = 5$), and '*Geographic Information System (GIS)*' ($f = 4$). This indicates that XR is adopting various methods to enhance technology-mediated accessibility. System elements, particularly '*Data Visualization*' ($f = 6$) and '*Data Sonification*' ($f = 2$), indicated a growing interest in multi-sensory output. Specific scenarios were another crucial area, covering diverse contexts, particularly in education, such as '*E-learning*' ($f = 8$) and '*Educational Settings*' ($f = 5$), as well as '*Telecommunication*' ($f = 4$), highlighting XR's role as a platform for remote MSE. Many of these topics focused on educational VR, indicating that XR online learning has the potential to overcome the barriers in FtF and informal learning [2]. System quality was also well-represented, with metrics such as '*Immersion*' ($f = 6$) underlining the importance of absorption in XR. Other relevant terms included '*User Experience*' ($f = 4$), '*Presence*' ($f = 4$), and '*Telepresence*', reflecting the focus on the continuity of personal identity within XR.

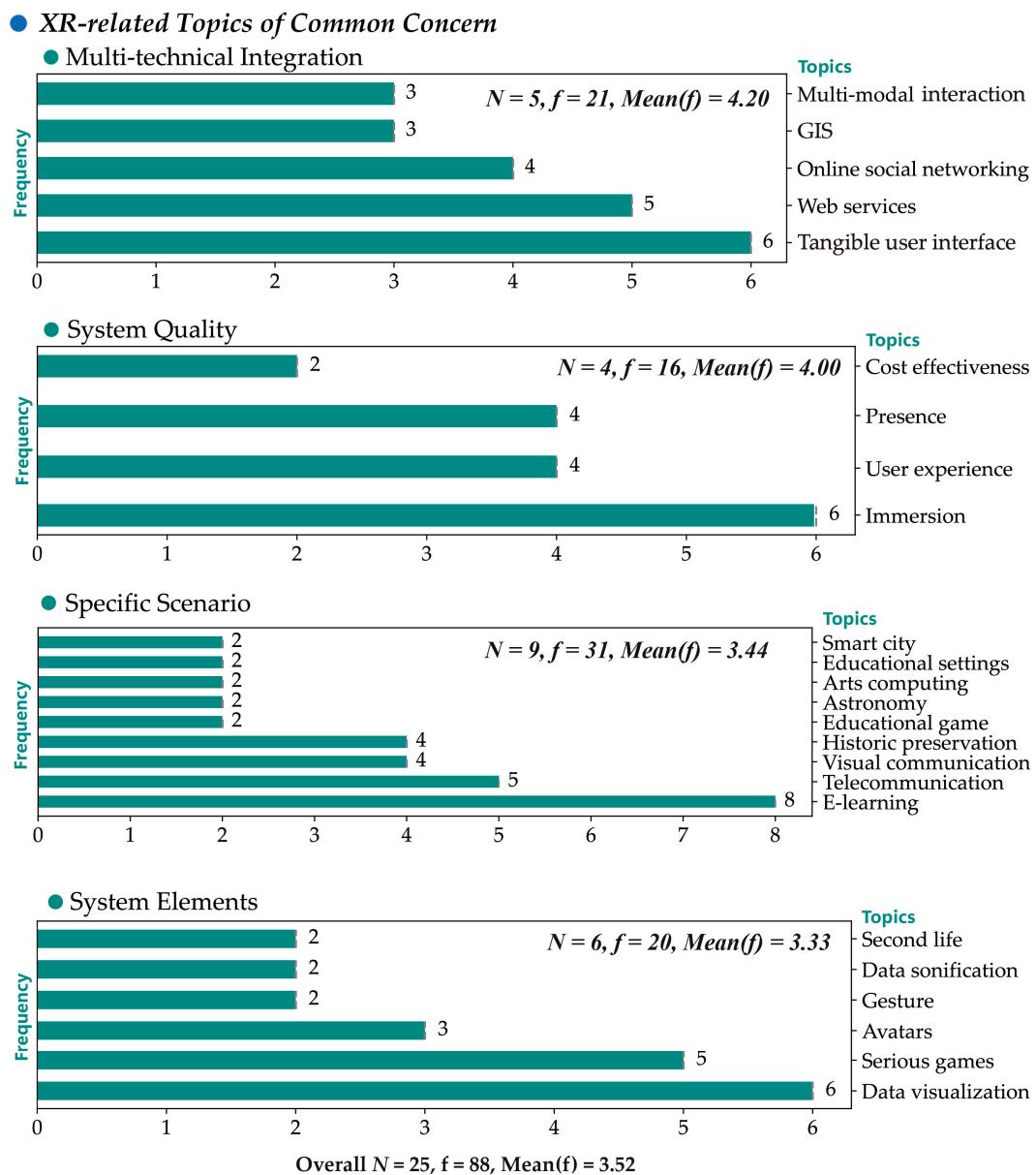


Figure 4. XR-related Topics in the Collected Samples.

In conclusion, a balanced yet nuanced focus on both MSE and XR-related topics is revealed. This comprehensive exploration highlights the early-stage richness of the field.

4.1.2. Application Domains, MSE Scales and XR Types

The samples span several domains: (1) Health Care and Rehabilitation, (2) Education and Training, (3) Business and Marketing, (4) Tourism and Hospitality, (5) Game and Entertainment, (6) Environment and Sustainability, (7) Culture Heritage and Exhibition, (8) Creativity, and (9) General Purpose. Education and Training emerged as the dominant domain (26/88, 29.5%), followed by Health Care and Rehabilitation (13/88, 14.8%). Environment and Sustainability and Cultural Heritage and Exhibition each accounted for 10 samples (10/88, 11.4%). However, Art (2/88, 2.3%) and Tourism (3/88, 3.4%) were less represented due to their nascent status over the past five years. Notably, there were 12 samples categorized as General Purpose (13.6%), indicating no application domain specified. This disparity suggests an evolving landscape with wide application scenarios, as well as the potential for new domains to surface and a more equitable distribution anticipated.

Three categories of MSE scales were identified within the samples: one-on-one engagement, group-level engagement, and public engagement. Group-level engagement was predominant, constituting 55.7% of cases (49/88), followed by public engagement at 29.5% (26/88) and one-on-one engagement at 14.8% (13/88). Temporal analysis indicated a significant increase in group-level engagement research over the past five years. Public engagement has shown fluctuating scholarly interest since 2005, with a recent resurgence likely due to technological advancements. In contrast, one-on-one engagement only became a notable area of study after 2017, with modest growth. These data reflect the scalability of MSE's core characteristic—'multi-users'—as evidenced by the high share of group-level and public engagement. Although one-on-one engagement has the smallest proportion, this dyadic form of interaction remains important. This dyadic form of interaction remains important, especially in areas such as games, entertainment, and general-purpose applications. It is mainly valued because it provides a preliminary testing environment that is less affected by external variables. It fosters strong connections where participants feel valued and focused, which can enhance overall social participation, increase trust, and build intimate relationships; thus, it deserves further attention.

Six categories of XR were identified: AR, MR, VR, VR+MR, VR+AR, and VR+MR+AR. Single-interface systems were predominant (92.0%, 81/88), with MR being the most common (29.5%, 26/88), followed by VR (36.4%, 32/88) and AR (26.1%, 23/88). Temporal trends reveal stable growth for MR and AR since 2002 and 2005, respectively. VR, introduced in 2011, experienced rapid expansion, particularly in the last five years (52.8%, 28/53). Multi-interface systems, though less frequent, showed an emerging preference for VR+MR (4.5%, 4/88), followed by VR+AR (2.3%, 2/88) and VR+MR+AR (1.1%, 1/88). Single-interface XR systems are evenly distributed and have significant growth in VR. Interest in multi-interface systems is developing and may increase as technology advances and research demands expand.

The correlations between application domains, XR Types, and MSE Scales were analyzed to understand their contextual relationships. Cramér's V¹ was used to calculate these correlations. The results indicated a moderately significant positive correlation between Application Domains and XR Types (Cramér's V = 0.360, $\chi^2(40) = 57.085$, $p = 0.039$) and between Application Domains and MSE Scales (Cramér's V = 0.564, $\chi^2(16) = 55.930$, $p = 0.000$). However, there was no significant correlation between XR Types and MSE Scales, thus excluded from further analysis. Figures 5 and 6 used alluvial maps to visualize these internal complex relationships.

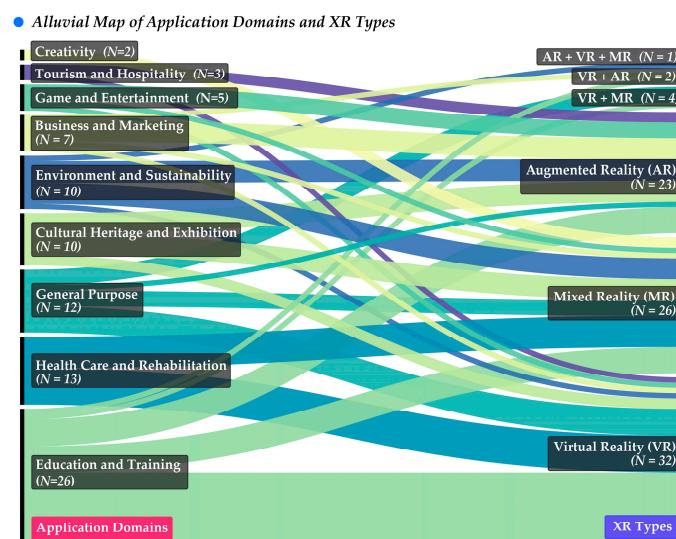


Figure 5. Alluvial Map of Application Domains and XR Types.

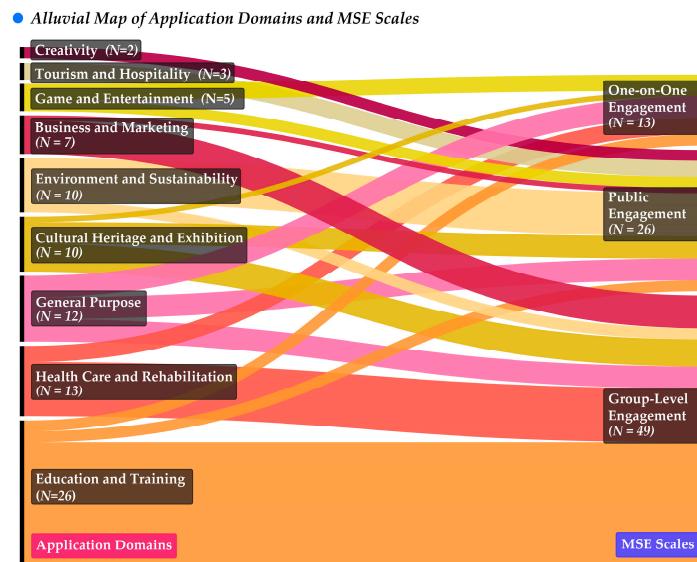


Figure 6. Alluvial Map of Application Domains and MSE Scales.

The predominant mapping between application domains and utilized XR Types was in Education and Training using VR, which accounted for 15.9% (14/88). This was followed by Health Care and Rehabilitation using VR at 8.0% (7/88) and Health Care and Rehabilitation using MR at 6.8% (6/88) (Figure 5). In the Cultural Heritage and Exhibition domain, the most common XR interfaces are based on AR and MR, with four cases each using them (4.6%, 4/88). In contrast, VR is less frequently applied, with only two cases involving it (2.3%, 2/88). In the General Purpose domain, the XR interfaces used, from most to least common, are VR (5.7%, 5/88), MR (3.4%, 3/88), and AR (1.1%, 1/88). This highlights that single-interface XR systems are widely applied across various domains, and their suitability for specific fields along with researchers' preferences for particular interfaces are becoming more evident. However, there remain unexplored areas, such as using single-interface XR systems in specific domains such as Health Care and Rehabilitation with AR and Creativity with VR. Multi-interface systems, being more recent developments, showed a limited presence across different domains.

The associations in Application Domains: MSE scales were led by Education and Training: Group-level Engagement (26.1%, 23/88) (Figure 6). This was followed by Health Care and Rehabilitation: Group-level Engagement (11.4%, 10/88), Environment and Sustainability: Public Engagement (9.1%, 8/88), and Business and Marketing: Group-level Engagement (6.8%, 6/88). Although these three mappings occur less frequently compared to Education and Training: Group-level Engagement, they are still widely referenced. Even fewer are the mappings for Cultural Heritage and Exhibition domain: Group-level Engagement (5.7%, 5/88). The above-shown mappings show associative solid tendencies. However, the General Purpose domain did not exhibit any significant preference across the different MSE scales. This is evident from its even co-occurrence across One-on-one, Public, and Group-Level Engagement, each accounting for 4.5% (4/88). When considering the mapping relationships between Application Domains and XR Types, it becomes clear that Education and Training, Health Care and Rehabilitation, and Environment and Sustainability are the three domains most strongly linked to XR Types and MSE Scales. In other words, these three domains are characterized by distinct technical support and audience attributes in XR-mediated MSE.

The extensive and widespread connections among modules indicate a clear tendency for technology and audience preferences in different XR-mediated domains. This highlights the gradual formation of distinctive research flows and focus. However, the unlinked modules should not be overlooked, as they indicate untapped areas and potential, particularly in emerging domains and various XR Types. For instance, there is a notable lack of empirical studies combining Creativity with AR.

4.2. The Mechanism of XR-Mediated MSE

4.2.1. Manifestation of Cognitive–Behavioral Mechanism

Frequency statistics revealed the tag distributions of MSE components (see Figure 7). In addition to the data distribution mentioned in the figure, several key manifestations of the cognitive–behavioral mechanisms merit further attention.

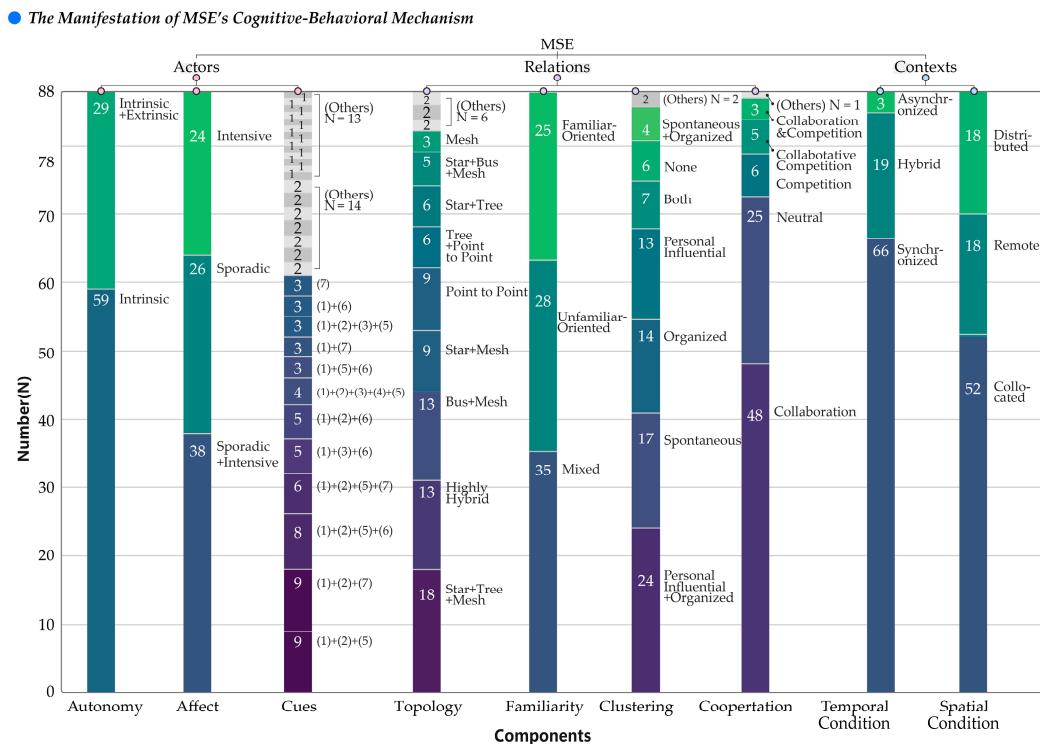


Figure 7. The Data Distribution of Components in MSE’s Cognitive–Behavioral Mechanism. Note: In the variable ‘Cues’, the following abbreviations are used: (1) signifies Verbal Communication, (2) signifies Facial Expression, (3) signifies Gesture, (4) signifies Touch, (5) signifies Posture, (6) signifies Object Control, (7) signifies Information Sharing. (2) Each Bar Representing the Components in This Figure Contains Sub-bars with a Gradient from Purple to Blue to Green, Indicating the Distribution of the Number of Manifestations Corresponding to the Components (e.g., Total N of Autonomy = 88, intrinsic N = 59).

For relation-related components, intrinsic motivation accounted for the majority of MSE activities at 67.0% (59/88). Examples include public exhibitions [161,170]. Activities involving a blend of intrinsic and extrinsic motivations, making up 33.0% (27/88), were notably more prevalent in structured environments. These include settings such as collaborative learning [143,147,182] and corporate innovation activities [121,137,150]. These blended motivations often supported group-level or goal-oriented engagements. In terms of ‘affect’, most activities induced mixed ‘sporadic + intensive’ states (43.2%). Sporadic affective states (29.5%, 26/88) were more frequently observed in spontaneous or public MSE activities, while purely intensive affective states (27.2%, 24/88) emerged in more structured and immersive experiences. Combinations of SI cues used in the XR environment varied and diversified. Notably, verbal cues were the main medium, used in 89.8% (79/88) of cases, while non-verbal modes were less common (10.2%, 9/88).

For relation-related components, a significant 81.8% (72/88) of relations used multiple topologies, highlighting the complexity of interpersonal relationships. Highly hybrid topology was typical in multi-stage or long-term engagements, such as the online ‘IEEE VR Conference 2021’ [175], which combined workshops, paper sessions, talks, panels, plenary sessions, demos, and posters. Another example is the role-playing game ‘Pokémon GO’, which integrated activities such as catching and trading Pokémons, training ‘Pokémon’, and

'Team GO Rocket social mechanisms' [109]. These diverse MSEs created various interpersonal topologies, resulting in the presence of highly hybrid topologies. In 'cooperation' models, pure collaboration dominated at 61.5% (48/78), with 25 instances lacking clear competition, i.e., neutral mode (28.4%). Pure competition and collaborative competition occurred in 6.8% (6/88) and 4.5% (4/88) of cases, while competitive collaboration was notably rare at 1.1% (1/88).

For context-related components, 'temporal conditions' in MSE showed a strong preference for synchronized communication, occurring in 75.0% (66/88) of instances. Purely asynchronous communication was less common, found in only 3.4% (3/88) of cases, though hybrid modes were noted in 19 samples (21.6%), indicating a growing recognition of asynchronous communication's benefits.

4.2.2. Manifestation of the Technology-Mediated Mechanism

Frequency statistics revealed the tag distributions of each SF (see Figure 8). In addition to the data distribution mentioned in the figure, several key manifestations of the technology-mediated mechanisms merit further attention.

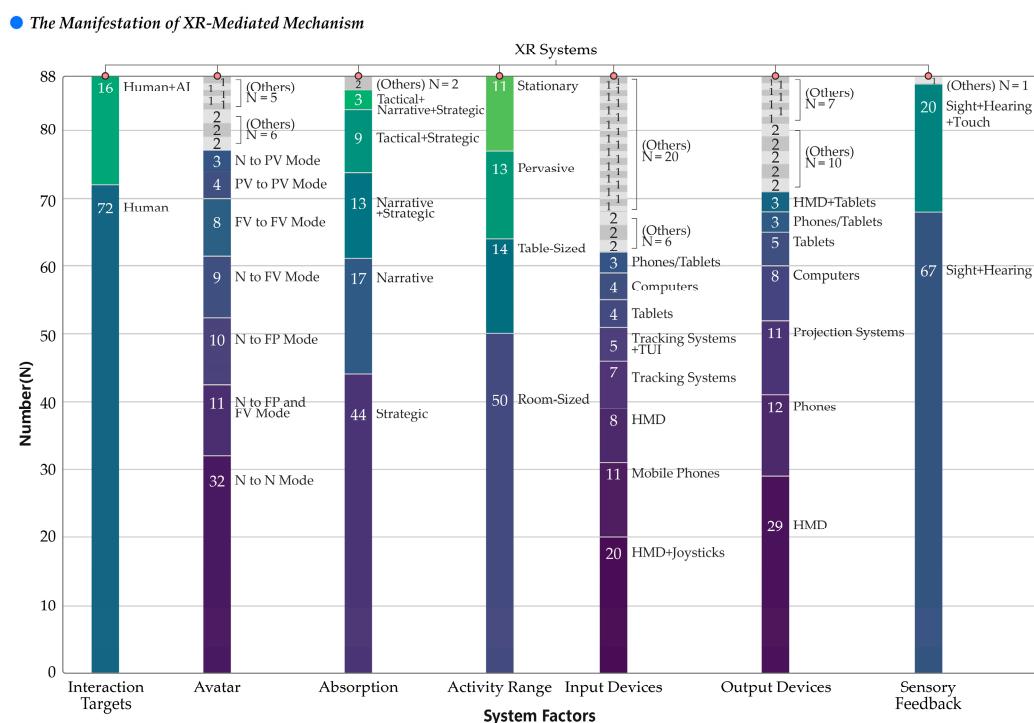


Figure 8. The Data Distribution of SFs in the XR-Mediated Mechanism. Note: (1) In the variable ‘Avatar’, the following abbreviations are used: N signifies Non-body, FV signifies Full Visual Body, FP signifies Full Physical Body, PP signifies Part Physical Body, and PV signifies Part Visual Body. (2) Each Bar Representing the SFs in This Figure Contains Sub-bars with a Gradient from Purple to Blue to Green, Indicating the Distribution of the Number of Manifestations Corresponding to the SFs (e.g., Total N of Interaction Targets = 88, Human N = 72).

Current MSE activities mainly involved human participants, accounting for 81.8% (72/88) of the cases. AI agents were involved in 18.2% (16/88) of interactions, with only two instances featuring physical AI, such as social robots [127,164].

In terms of ‘avatar’ settings, the noteworthy data distributions are as follows: (1) Users without avatars, observing others’ full physical and virtual bodies from a first-person perspective (12.5%, 11/88). (2) Users without avatars, observing others’ full physical bodies from a first-person perspective (11.4%, 10/88). (3) Users without avatars, observing others’ full virtual bodies from a first-person perspective (10.2%, 9/88). Sometimes, the full virtual bodies are AI-driven; this can be understood as AI being designed as a user’s proxy,

representing them in virtual environments. For example, a virtual assistant can interact with other users or systems on behalf of the user when they are not present, effectively serving as an avatar. This indicates a diverse use of avatars, with significant variation in visual representations and representational symmetry.

‘Absorption’, was primarily characterized by strategic immersion, which was observed in 50.0% (44/88) of the cases. Examples of such immersions included goal-directed activities such as modules of inspirational stimuli [110,115,121,137], simulation-based learning [106,128,159,177,182], and engaging gameplay [114,122,157,174]. This far exceeded narrative immersion at 19.3% (17/88) and combined narrative and strategic immersion at 14.8% (13/88). However, tactical immersion was less common, being addressed in only 12 samples. It was typically combined with strategic and narrative forms of immersion (13.6%), often appearing in walkable physical exercises, such as co-participatory planning in actual walkable environments [118,130].

In terms of ‘sensory feedback’, the majority of users experienced visual-auditory feedback, which was present in 76.1% (67/88) of the cases. Tactile feedback was noted in 22.7% (20/88) of the interactions. Olfactory feedback was nearly non-existent, appearing in only 1.1% (1/88) of cases, and there was no recorded use of gustatory feedback.

4.2.3. Strongly Correlated SFs for Mediating MSE

The XR-mediated influence on MSE was examined through a correlation analysis. A heat map was used to visualize potential variable group combinations (i.e., ‘SFs → Components’) (see Figure 9). Correlation analysis indicates that all potential SFs influence at least one component of the cognitive-behavioral mechanism, with correlations (Cramér’s V > 0.200) and statistical significance ($p < 0.05$). Moderate correlations (0.200 < Cramér’s V < 0.600, $N = 10$) to high correlations (Cramér’s V > 0.60, $N = 3$) were observed. Among these, eight strong correlations were identified with positive effects from SFs. Additionally, two strong correlations were found with negative effects from certain components. Other correlations were excluded due to being counterintuitive. For detailed co-occurrence tags, refer to Supplementary File S3.

● The Correlation Analysis between SFs and Components

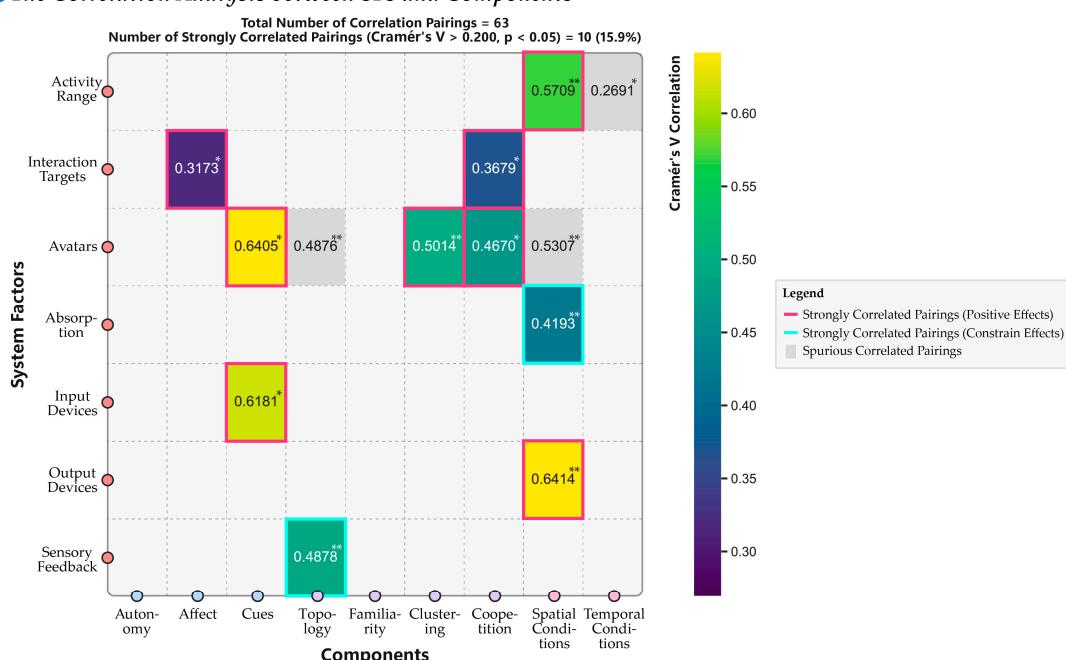


Figure 9. The Heatmap of the Correlations between Components of the MSE and SFs of XR. Note: Note: The Correlation between the Variables was Indicated by “**” to Signify Its Significance at the 0.05 Level, Namely $p < 0.05$, “***” to Signify its Significance at the 0.01 Level, Namely $p < 0.01$.

- (1) Output devices as strongly correlated SF for spatial conditions. Output devices mediate spatial conditions based on significant correlations (Cramér's V = 0.641, $\chi^2(40) = 72.403, p = 0.001$). For instance, projection + speaker setups are exclusive to collocated conditions, creating dense, immersive communication environments for attendees. Meanwhile, Samayoa et al.'s projection system [123] also enhanced social presence in remote conditions. HMD and phones are more associated with collocated conditions, while computers are linked with remote conditions. Overall, corresponding output devices mediate spatial conditions; further case analyses are provided in Section 5.2.3 Output Devices → Spatial Conditions;
- (2) Input devices as strongly correlated SF for cues. Input devices mediate cues with significant correlations (Cramér's V = 0.618, $\chi^2(930) = 1008.478, p = 0.037$). Examples include HMD+Joystick and various cue expression paradigms, tracking systems involving postures, and the amplification of body communication in MSE. TUI introduces object control cues, showing their potential as communication tools, as seen in cases [58,129,146,149,152,156,171]. Further enhanced modes are provided in Section 5.2.3 Input Devices → Cues;
- (3) Avatar as strongly correlated with SF for cues, clustering, and coopetition. The [Avatar → Cues] pairing showed a highly significant correlation (Cramér's V = 0.641, $\chi^2(434) = 505.415, p = 0.010$). Further case analyses are provided in Section 5.2.3 Avatars → Cues. The [Avatar → Clustering] pairing also showed a significant correlation (Cramér's V = 0.501, $\chi^2(84) = 123.675, p = 0.003$), with avatars facilitating organized and personal clustering. This suggests that avatars may facilitate personal Influential clustering by extending real-world identities. It provides anonymity or enables identity reshaping in massive social settings, granting more freedom for social expression, as seen in the SocialVR application [112] and VRdeo platform [181]. Further case analyses are provided in Section 5.2.3 Avatars → Coopetition. For [Avatar → Coopetition], a significant moderate correlation was found (Cramér's V = 0.467, $\chi^2(70) = 95.964, p = 0.021$), with numerous tags binding its presence to collaboration. This proves its potential in supporting virtual collaboration, similar to non-avatar modes, as in cases [46,51,177,181,189]. Despite moderate significant correlations for [Avatar → Topology] and [Avatar → Spatial conditions] (Cramér's V = 0.488, $\chi^2(154) = 230.184, p = 0.000$; Cramér's V = 0.531, $\chi^2(28) = 49.563, p = 0.007$). These correlations are based on fragmented, low-co-occurrence tag combinations, indicating a pseudo-correlation. These tags' low-frequency co-occurrences do not support causality and contradict logical expectations. Therefore, avatars do not mediate topology and spatial conditions. Further case analyses are provided in Section 5.2.3 Avatars → Clustering;
- (4) Activity range as strongly correlated SF for spatial conditions. Activity range mediates spatial conditions based on significant correlations. The pairing [Activity range → Spatial Conditions] showed a significant moderate correlation (Cramér's V = 0.571, $\chi^2(6) = 57.358, p = 0.000$). Room-sized and table-sized ranges are more suited to collocated settings, pervasive ranges to hybrid settings, and stationary ranges to remote settings. However, designing static activity ranges for spatial conditions may limit designers' creativity. Further analysis is provided in Section 5.2.3 Activity Range → Spatial Conditions. Activity range and temporal condition also showed significant moderate correlations (Cramér's V = 0.269, $\chi^2(6) = 12.742, p = 0.047$). However, this only indicates a co-occurrence relationship, not causality. The correlation and significance levels are primarily due to synchronized interactions in room-sized ranges in many samples. Room-sized activity ranges inherently require less asynchronous and mixed communication, favoring immediate and cohesive interactions. For example, in the case of XRPublicSpectator [190], users in a room-sized range watched presenters engage in a high-interaction-density card battle game, negating the need for asynchronous or mixed communication channels. Other representative cases

- include [125,180,183]. Thus, activity range cannot be considered a strongly correlated SF for mediating temporal conditions;
- (5) Interaction targets as strongly correlated SF for affect and coopetition. Interaction targets are strongly correlated with affect and coopetition. Interaction targets mediate both affect and coopetition based on significant correlations. A moderately significant correlation was found between interaction targets and affect (Cramér's V = 0.317, $\chi^2(2) = 8.861, p = 0.012$). Further case analyses are provided in Section 5.2.3 Interaction Targets → Affects. For coopetition, a moderately significant correlation (Cramér's V = 0.368, $\chi^2(5) = 11.91, p = 0.036$) indicates AI's potential in mediating collaborative settings. Thus, interaction targets and coopetition are strongly correlated. Further case analyses are provided in Section 5.2.3 Interaction Targets → Coopetition;
 - (6) Spatial conditions as strongly correlated with SF for absorption. Absorption mediates spatial conditions based on significant correlations. The [Absorption → Spatial Conditions] pairing showed moderate positive correlations (Cramér's V = 0.419, $\chi^2(10) = 30.946, p = 0.001$). Further case analyses are provided in Section 5.2.3 Spatial Conditions → Absorption;
 - (7) Topology as strongly correlated component for limiting the SF sensory feedback. Topology mediates sensory feedback based on significant correlations, with a moderately significant correlation (Cramér's V = 0.488, $\chi^2(22) = 41.872, p = 0.007$). Topology constrains the achievable sensory feedback, and the desired MSE structure limits technological deployment, indicating a reciprocal influence (see Section 5.2.3 Topology → Sensory Feedback).

4.3. The Effectiveness of XR-Mediated MSE

4.3.1. Settings of Evaluation

Basic Settings

In examining the overall deployment of XR-mediated MSE evaluation, several dimensions were analyzed based on statistics of tagging data.

- (1) Goals. Assessing the comprehensive effect of XR-mediated MSE is becoming increasingly significant, although many studies still focus on evaluating either MSE or XR system aspects separately (Figure 10). MSE-related goals primarily explore preferences in MSE activities, such as perceptions, engagement, and task efficacy. System-related goals focus on technical effectiveness, including usability, immersion, and interface design. Among the 88 cases, 84 involved MSE-related goals (95.4%), compared to only four focusing solely on testing the XR system (4.5%);

• Analysis of Goals, Types, and Durations

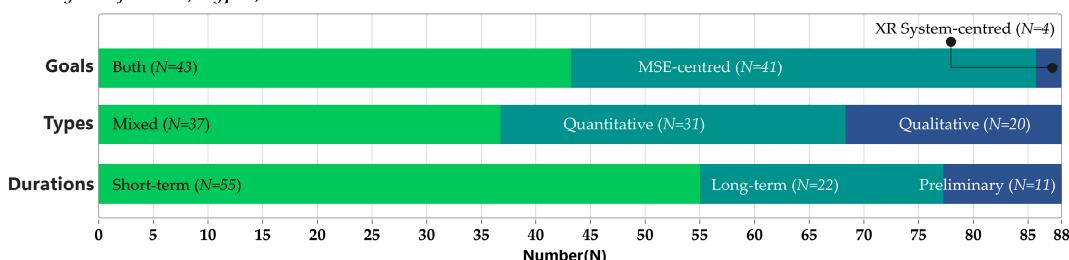


Figure 10. Analysis of Goals, Types, and Durations. Note: The light green, medium blue-green, and dark blue colors in the figure represent the number of the analyzed items' (e.g., goals) samples ranked first, second, and third, respectively.

- (2) Types. There is a clear preference for mixed (combined qualitative and quantitative) and purely quantitative research types (Figure 10). The data show that mixed methods were used in 37 samples (42.0%), purely quantitative research in 31 samples (35.2%), and purely qualitative types in 20 samples (22.7%);

- (3) Durations. Study durations varied significantly to meet diverse research requirements, with a tendency towards short-term evaluations (Figure 10). Preliminary evaluations, aimed at quickly gathering initial user feedback to refine early prototypes, constituted 12.5% (11/88) of the samples. Short-term studies involving advanced XR systems employed formal, intensive experimental designs, with 62.5% (55/88) lasting a month or less, primarily conducted in field and laboratory settings. Only 22 samples (25.0%) extended beyond one month, typically featuring multiple experimental deployments to assess sustainability and potential long-term effects of XR-mediated MSE.
- (4) Size. The size of participants varied widely across studies, with a noticeable trend toward increased average engagement in evaluations (Figure 11). Five samples did not report explicit participant numbers, comprising 5.7% (5/88) of the total. The trendline ($y = 2.23x - 4411.93$) created by linear regression was added programmatically to visually aid the understanding of variability and dispersion in participant numbers [191]. It is crucial to highlight two aspects. First, the trendline appeared less pronounced due to the logarithmic scale on the vertical axis. Second, the extremely low coefficient of determination ($R^2 = 0.003$) suggested a substantial dispersion of data points, underscoring significant variability in participant numbers across samples. Specifically, 83 samples showed that, on average, around 88 individuals participated in each case, with considerable variability observed (mean = 87.80, standard deviation = 181.84);

• Analysis of Participant Size

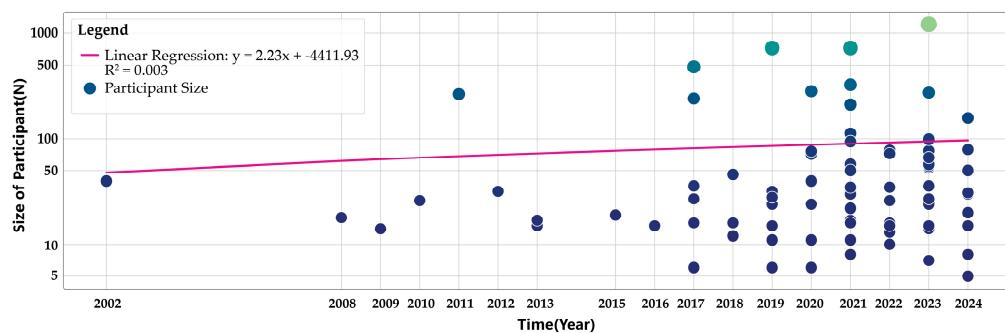


Figure 11. Analysis of Participant Size.

- (5) Methods. The samples revealed significant diversity in organizing research methods in XR-mediated MSE studies (see Figure 12). Multiple methods were often employed within each sample, totaling 246 recorded uses. Questionnaires emerged as the most common method, constituting 28.5% (70/246) of all methods utilized, and were pivotal in collecting standardized feedback across studies. Interviews, accounting for 12.2% (30/246), were crucial for gaining deep insights. Observations, representing 11.0% (27/246), provided real-time insights into user behaviors and interactions. Workshops, which facilitated co-creation in evaluation settings, were used in 8.9% (22/246) of instances. Field experiments and control experiments were used with similar frequency, being employed 20 times (8.1%) and 19 times (7.7%), respectively. Video coding and the emerging method of ‘virtual field trips’ (VFT), each making up 5.7% (14/246), underscored their importance in behavioral analysis and immersive evaluations. VFT has gained popularity for simulating real-life field experiences in scenarios where onsite visits are impractical. Community-Based Participatory Research (CBPR), though less common at 3.7% (9/246), emphasizes the collaborative involvement of researchers, government, and community members throughout the research process. It focuses on joint decision-making, social justice, and community empowerment. This approach is typically found in studies related to the environment and sustainability, as seen in representative cases [124,130,163,171,176]. In contrast, workshops are more short-term and researcher-led, with a product/service improvement orientation. They focus on obtaining user perspectives rather than empowerment and are usually found in studies related to education and business, as seen in representative

cases [116,120,129,140,155,172,175,185]. Physiological measurements and self-reports were rarely employed, appearing just twice in the study (0.8%, 2/248). Only two studies incorporating physiological sensors to assess bodily, cognitive, and social responses—namely ECG (electrocardiography) and EDA (electrodermal activity) in Sayis et al. [156] and EEG (electroencephalogram) in Fan et al. [164]. Behavioral indicators such as control data, task status, performance data, and state anxiety could be identified in these studies, representing 0.8% (2/248). Physiological measurements, which deliver precise, objective data, could be limited by their complexity and associated high costs. Two studies utilized self-reports to collect data. One study combined self-reports with video coding, psychophysiological data, and system logs to comprehensively assess users' observable SI behaviors [156]. The other study focused on students' self-reported feelings about a lab activity. Compared to structured questionnaires, self-reports are more qualitative, allowing participants to provide detailed and personalized responses [151]. This approach captures deeper personal insights and richer social context information. However, self-reports may be influenced by participants' memory biases and inconsistencies in measurement dimensions, which limits their use.

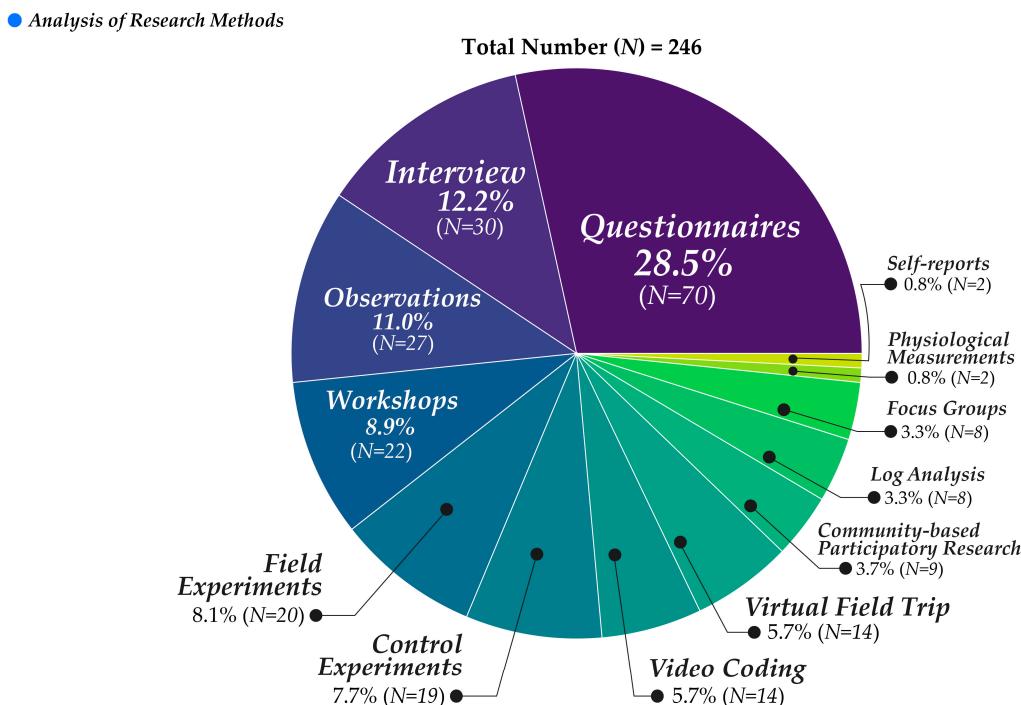


Figure 12. Analysis of Research Methods. Note: The sectors in the pie chart do not add up to 100%, but rather 99.7%, due to a very slight rounding difference. This minor discrepancy does not affect the accuracy of the arguments and conclusions in the main text.

Metric Settings

The evaluation metrics were recorded and analyzed to understand the specific MSE-related and XR-related evaluation content and their differences. The data sources comprised the scales and metrics used in the samples. Notably, synthesis and deduplication of the data were conducted. A total of 32 samples (36.4%, 32/88) used scales, resulting in 68 instances of scale application. Before 2017, only three studies employed scales (3.41%, 3/88); their usage significantly increased post-2017. These scales fell into two categories (see Tables 1 and 2).

Table 1. Analysis of the used MSE-related scales in the samples.

Used MSE-Related Scales in Samples			Time
Scales with Frequency > 1	Frequency (f)	Reference	
<i>Group Environment Questionnaire (GEQ)</i>	5	[46,144,149,155,167]	2017
<i>Self-Assessment Manikin (SAM)</i>	3	[144,167,177]	2017
<i>Networked Minds Measure of Social Presence Inventory (NMMSPI)</i>	2	[122,155]	2008
<i>Montreal Cognitive Assessment (MoCA)</i>	2	[164,177]	2022

Scales with Frequency = 1: *Fugl-Meyer Assessment of the Upper Extremity (FMA-UE)*, *User Engagement Scale (UES)*, *Visual and Cognitive Heuristics*, *Ethographic and Laban-Inspired Coding System of Engagement (ELICSE)*, *Motivated Strategies for Learning Questionnaire (MSLQ)*, *Happiness Scale*, *Topographic Map Assessment (TMA)*, *Slater, Usoh, and Steed Questionnaire*, *The Affective Slider (AS)*, *State Trait Anxiety Inventory for Children (STAIC)*, *Person with Dementia Scale (EPWDS)*, *Observed Emotional Rating Scale (OERS)*, *Observational Measurement of Engagement (OME)*, *Memorable Tourism Experience (MTE)*, *Mini-mental State Examination (MMSE)*, *Montreal Cognitive Assessment (MSA)*, *Positive and Negative Affect Schedule (PANAS)*, *20-Item Short Form Health Survey (SF-20)*, *Multidimensional Mood State Questionnaire (MMSQ)*, *Acceptance of Head-mounted Virtual Reality in Older Adults*, *Social Presence Scale (SPS)*, *Willingness and Likeliness to Reconnect Scale*, *Affect Grid (AGRID)*, *Social Presence Scale (SPRES)*, *Collaboration Self-Assessment Tool (CSAT)*, *Intrinsic Motivation Index (IMI)*, *Teamwork Satisfaction Questionnaire*, *User Experience in Immersive Virtual Environments*, *Collaborative Learning in VR for Cross-disciplinary Distributed Student Teams*, *Virtual Embodiment Questionnaire (VEQ)*.

Table 2. Analysis of the used XR-related scales in the samples.

Used XR-Related Scales in Samples			Time
Scales with Frequency > 1	Frequency (f)	Reference	
<i>System Usability Scale (SUS)</i>	9	[46,118,123,141,144,152,157,181,183]	2013
<i>Simulator Sickness Questionnaire (SSQ)</i>	5	[46,51,134,177,183]	2023
<i>NASA Task Load Index (NASA)</i>	3	[134,177,183]	2023
<i>MEC Spatial Questionnaire (MEC)</i>	2	[134,177]	2023

Scales with Frequency = 1: *Usability Heuristics*, *Post-Study System Usability Questionnaire (PSSUQ)*, *Usability Metric*, *Computer Self-efficacy Scale (CSS)*, *Mobile Device Proficiency Questionnaire (MDPQ)*, *Simulation Task Load Index (SIM-TLX)*, *Slater–Usoh–Steed*, *Multimodal Presence Scale*.

- (1) MSE-related scales. The *Group Environment Questionnaire (GEQ)* was frequently used (7.4%, 5/68), and the *Self-Assessment Manikin (SAM)* was used less often (4.4%, 3/68), both starting in 2017 (see Table 1). This trend indicates a growing recognition of the importance of measuring group dynamics and emotional responses. The *Networked Minds Measure of Social Presence Inventory (NMMSPI)* and the *Montreal Cognitive Assessment (MoCA)* were also repeatedly utilized (each 2.9%, 2/68), focusing on social presence and cognitive ability, respectively;
- (2) XR-related scales. The *System Usability Scale (SUS)* was the most commonly used (13.2%, 9/68) (see Table 2). The *Simulator Sickness Questionnaire (SSQ)* (7.4%, 5/68) and the *NASA Task Load Index (NASA)* (4.4%, 3/68) were more frequently used than the *MEC Spatial Presence Questionnaire (MEC)* (2.9%, 2/68). The consistent use of the *SUS* since 2013 underscores a strong consensus on its effectiveness in evaluating user interface ease and efficiency in XR technologies. The application of the *MEC* highlighted attention to spatial presence driven by technological quality. The use of the *NASA* and *SSQ* scales, beginning in 2023, indicates an emerging interest in addressing cognitive load and physical discomfort caused by system tasks and hardware. Despite their lower overall frequency, XR system-related scales demonstrated more consensus in their application compared to MSE-related scales. This indicates that while XR scales are less commonly used, they achieve greater agreement among researchers. In contrast, MSE-related scales were more varied, with many being used only once.

The diversity in MSE-related and XR system-related metrics illustrates the multifaceted nature of evaluation contexts and the ongoing development of assessment frameworks. Each sample's metrics were documented and analyzed based on Social Network Analysis (SNA). SNA, based on *Graph Theory* [57], is a statistical method used to monitor, measure, and evaluate information flow. It transforms ambiguous knowledge into clear patterns and processes through multiple dimensions of knowledge resources and strict methodologies. This makes it suitable for analyzing the evaluation metrics of XR and MSE, as well as their co-occurrence relationships in this study [192]. The metric network of MSE-related and XR system-related metrics was visualized using the algorithm *Eccentricity*², which measures the greatest distance from a node to any other node in the network [193]. Metrics were presented as nodes within the network. Nodes closer to the center of the network had lower eccentricity, indicating they were more frequently adopted as indicators by samples. The thickness of the edges between the nodes represented the higher frequency of co-usage. MSE-related and XR-related metric networks are detailed in Figures 13 and 14, respectively.

● MSE-related Metric Network

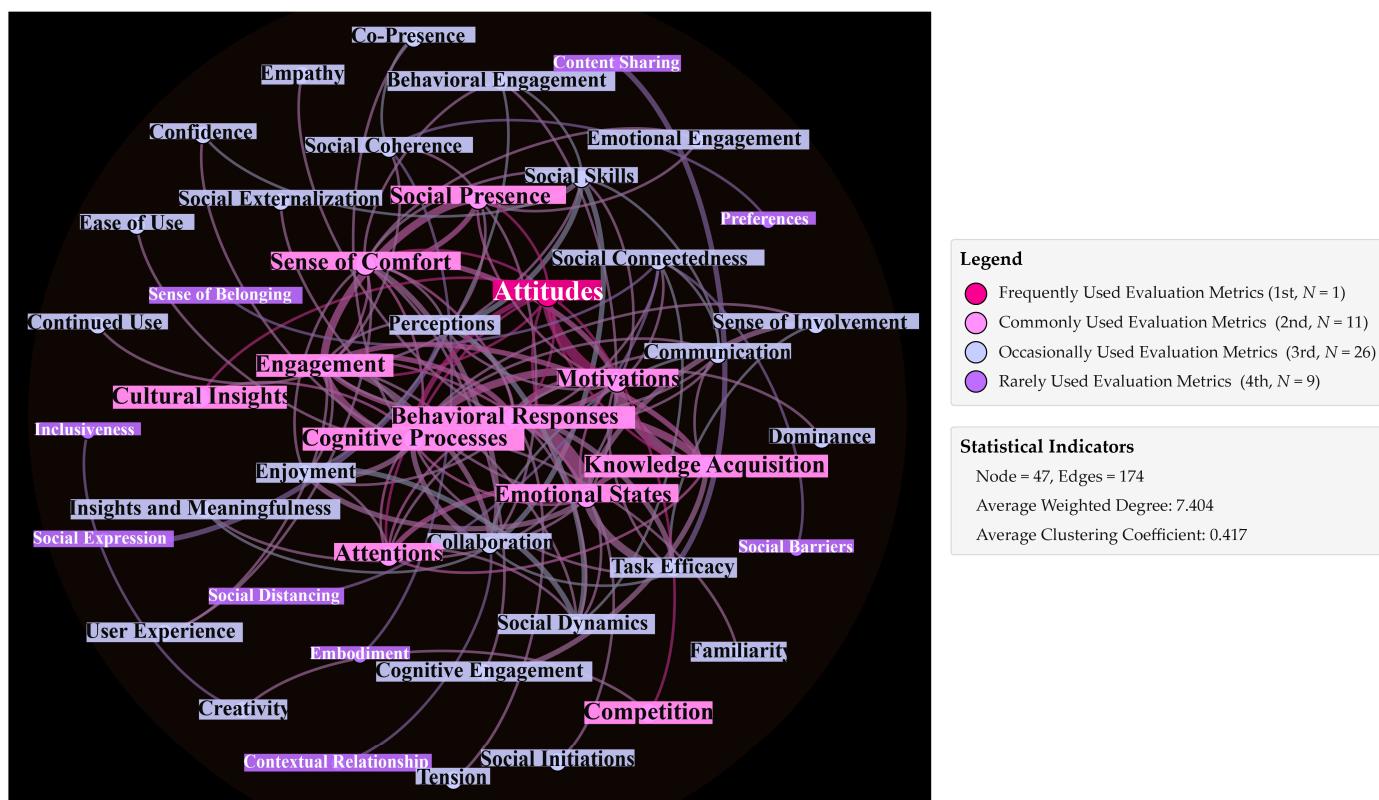


Figure 13. MSE-related Metric Network.

- (1) MSE-related metric network. This network was relatively dispersed, with 47 metrics and 174 connections, an average weight degree of 7.404, and an average clustering coefficient of 0.417. The frequent metric was 'Attitude', reflecting subjective feedback on engagement (2.1%, 1/47). Secondary metrics with high co-occurrence included 'Motivations', 'Emotional State', 'Cognitive Process', and 'Social Presence', covering various engagement stages (23.4%, 11/47). Tertiary metrics (55.3%, 26/47), used occasionally, offered finer content granularity and stemmed from secondary metrics, such as 'Perceptions' from 'Motivations' and 'Creativity' from 'Cognitive Process.' Quaternary metrics (19.1%, 9/47), rarely used, evolved from tertiary metrics, and had very specialized characteristics, such as 'Social Barriers' and 'Social Distancing';
- (2) System-related metric network. This network was denser, with 26 metrics and 67 connections, an average weight degree of 5.154, and an average clustering coefficient of

0.481. Here, '*Usability*' stood out as the sole primary metric (3.8%, 1/26), emphasizing its authoritative role in evaluations. Secondary metrics (69.2%, 18/26), often used in conjunction with '*Usability*', reflected distinct XR system features such as '*Immersion*', '*Presence*', and '*Spatial Presence*'. Tertiary metrics (23.1%, 6/26) included niche ones such as '*Plausibility Illusion*' and '*Body-ownership*'.

By comparing the average weight degree and number of metrics between MSE-related and XR-related metric networks. The MSE-related network has higher values, yet the XR-related network exhibits a higher average clustering coefficient. This indicates that MSE-related metrics are more diverse and less densely connected, resulting in a more dispersed and heterogeneous network. Specifically, the MSE-related network has a larger number of marginalized, low-connected nodes compared to the XR-related network. Conversely, the XR-related network is denser, with more tightly connected nodes and fewer low-connected nodes. This suggests that researchers have reached a greater consensus on XR-related metrics, which have formed a relatively well-established system of use. In contrast, MSE-related metrics are still evolving, with new metrics continuously being added. This is partly due to the complex tasks and contexts associated with MSE. Furthermore, it reflects that the evaluation criteria for XR-related metrics are more convergent and objective, while those for MSE-related metrics are more divergent and subjective.

● XR-related Metric Network

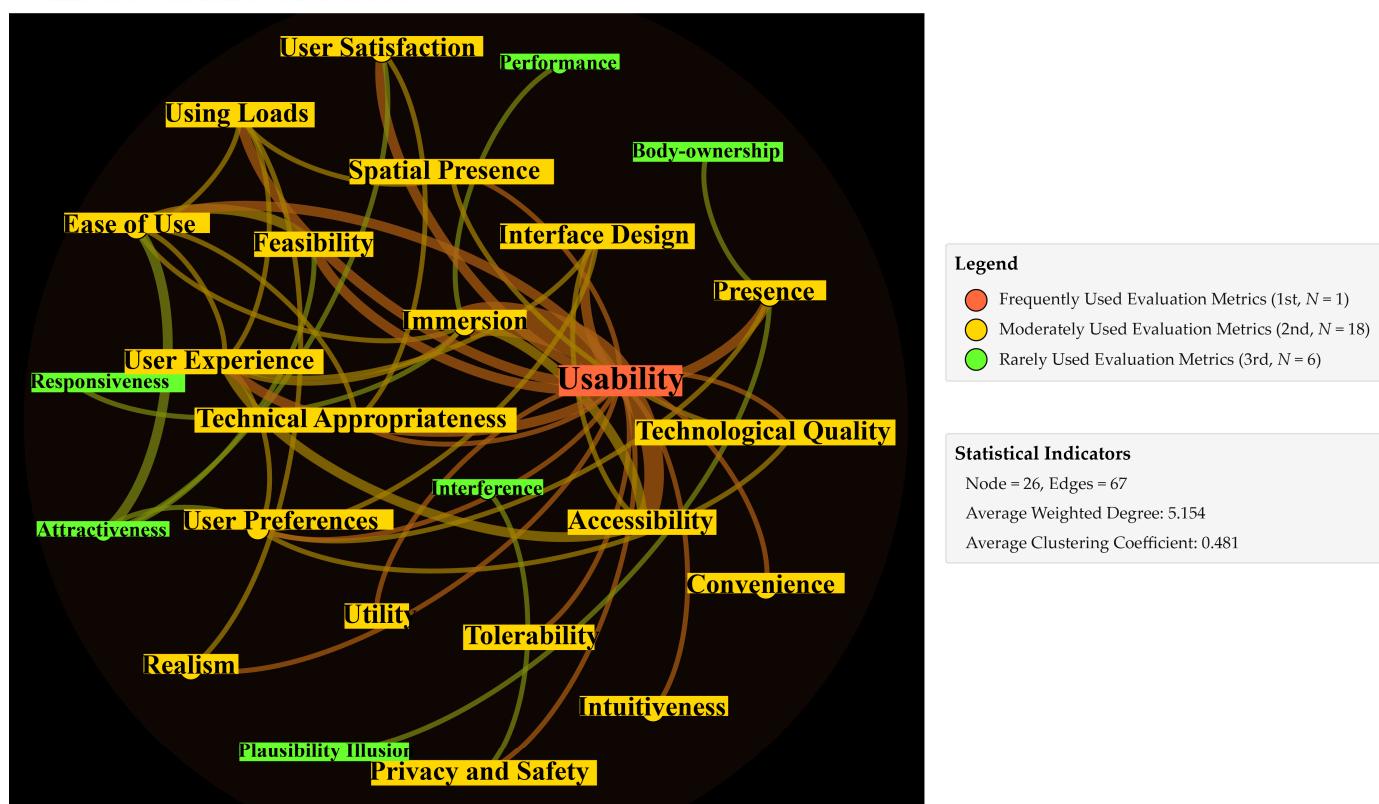


Figure 14. XR-related Metric Network.

The evaluation of XR-mediated MSE research reveals a complex and multifaceted experimental landscape, particularly in highlighting the widespread use of mixed methods and diverse research goals, methodologies, and durations. These studies demonstrate a profound focus on technology, cost-effectiveness, and user experience, progressively shaping a mature evaluation framework.

4.3.2. Performances of Evaluation

To assess the effectiveness of current XR-mediated MSE, both MSE-related and XR-related evaluation outcomes of samples were collected, drawing on positive (see Figures 15 and 16) and negative effects (see Figures 17 and 18). Notably, Figure 15 uses blue, purple, and pink lines to represent effects from the different components of the XR-mediated MSE analysis framework introduced in Section 2.

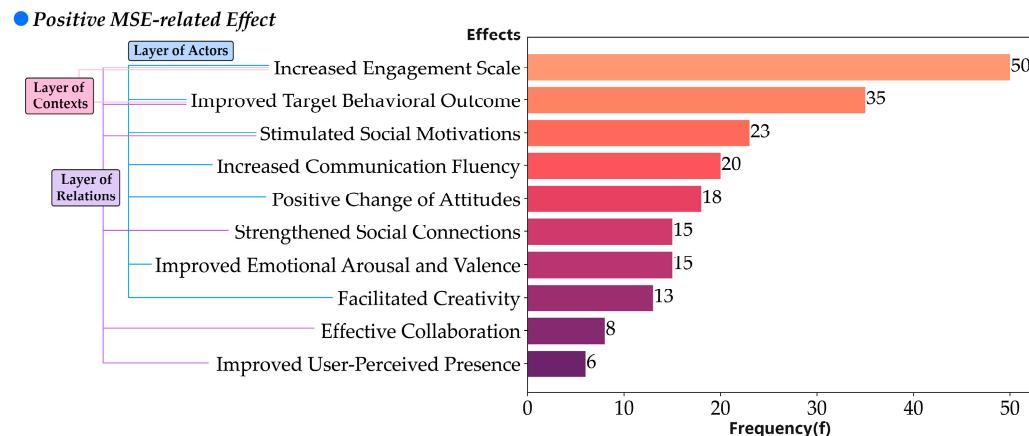


Figure 15. Positive MSE-related Effect.

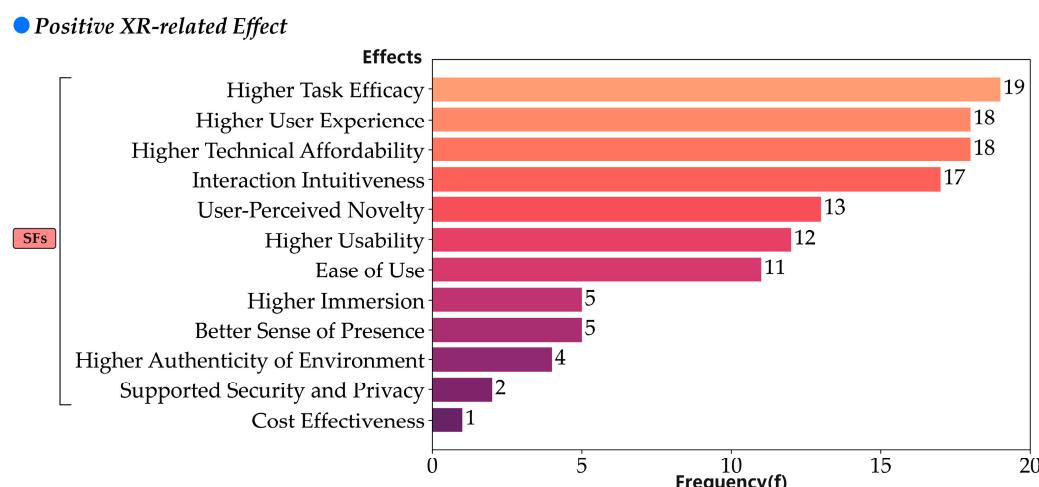


Figure 16. Positive XR-related Effect.

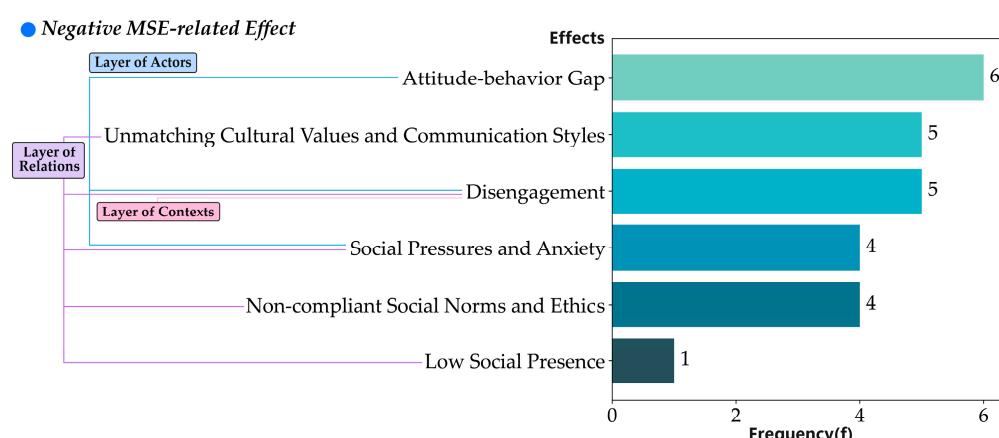


Figure 17. Negative MSE-related Effect.

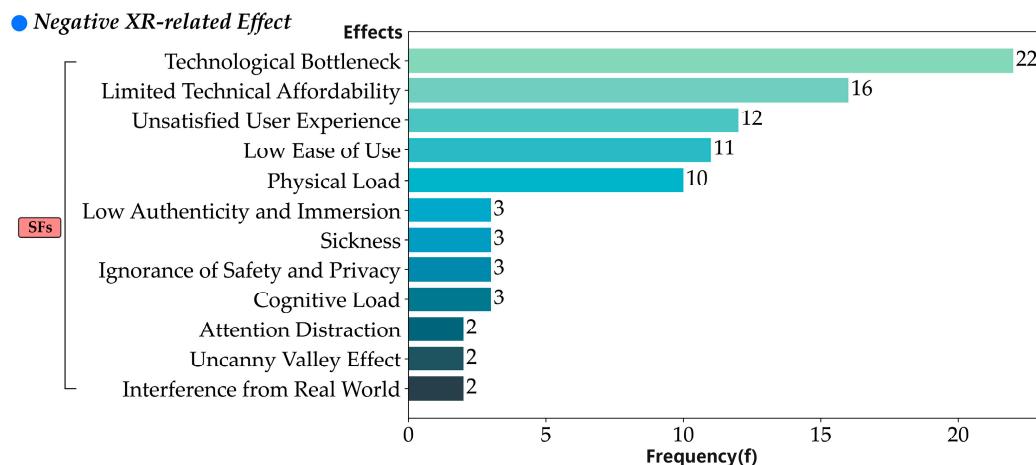


Figure 18. Negative XR-related Effect.

Positive Effects

- (1) MSE-related effects. The most significant positive outcome is the '*Increased Engagement Scale*' ($f = 50$), driven by the expanded scalability of participant groups. Another notable metric is '*Improved Target Behavioral Outcomes*' ($f = 35$), frequently observed in educational settings, where it improves learning outcomes, understanding, and reflection abilities among learners, and generates cultural insights in the cultural heritage domain. At the actor layer, '*Increased Communication Fluency*' ($f = 20$) is particularly prominent in MR environments due to the potential for natural interactions, such as gesture-based interactions [23,164], TUI [129,152], and tracking systems [127,161]. These systems allow users to focus more on expressing cues, especially verbal and posture cues, with greater physical freedom (further explained in Section 5.2.3 Input Devices → Cues). Another significant effect is the '*Positive Change of Attitudes*' ($f = 18$), which is more common in public MSE [124,138,163]. For example, Yavo-Ayalon et al. [124] used technology to help groups form bottom-up, informal, and low-mediation collaborative relationships, fostering a sense of community responsibility among members. Additionally, '*Improved Emotional Arousal and Valence*' ($f = 15$) is widely noted, especially in group-level MSE, where strong connections in tree/star topologies provide favorable conditions for emotional shifts (see cases: [158,173,177,179]). At the relational layer, '*Effective Collaboration*' ($f = 8$) and '*Improved User-Perceived Presence*' ($f = 6$) are primarily attributed to avatar-based interactions, which highlight the importance of avatars for social presence [51,132,134,177] (further cases in Section 5.2.3 Avatars → Cues);
- (2) XR-related effects. '*Higher Task Efficacy*' ($f = 19$) emerges as the most frequently reported positive effect, referring to enhanced performance in various MSE tasks across three primary domains: in education (e.g., addressing cross-cultural stereotypes [186]), rehabilitation (e.g., upper limb and head flexibility training [164]), and general purposes (e.g., gesture-based virtual object manipulation [183]). '*Higher Usability*' ($f = 12$) and '*Ease of Use*' ($f = 11$), while conceptually similar are and are both widely mentioned. Usability focuses on the effectiveness of new interaction mechanisms and system performance. In contrast, ease of use is typically measured by error rates and task completion times, especially in complex systems. This highlights the positive mediating role of input devices on social interaction cues (further explained in Section 5.2.3 Input Devices → Cues). '*Higher Immersion*' ($f = 5$), '*Better Sense of Presence*' ($f = 5$), and '*Higher Authenticity of Environment*' ($f = 4$) are more closely associated with VR or VR-based multi-interface systems. For example, cases [110–112,114,119,182] indicate higher immersion, primarily stemming from image rendering and audio-visual output quality. Cases [46,144,189] confirm enhanced sense of presence, defined as perceptions of virtual body ownership and consistency of one's sense of self with the

virtual environment. These three aspects are all related to spatial settings, underscoring the positive mediating role of absorption on spatial conditions (further detailed in Section 5.2.3 Interaction Targets → Affects). Lastly, ‘Cost Effectiveness’ (f = 1) is exemplified by Koukopoulos and Koukopoulos’s ‘Active Visitor’ system, which developed an asynchronous AR library application, expanding the scale and fluidity of visitor engagement and achieving more effective profit gains within established activity range limitations [113] (further detailed in Section 5.2.3 Activity Range → Spatial Conditions).

Negative Effects

- (1) MSE-related effects. ‘Disengagement’ (f = 5) is a multifaceted negative outcome, referring to low participation or withdrawal. This issue in public MSE primarily arises from unequal cooperative relationships due to imbalanced social topology (see Section 5.3). Distinct from ‘Disengagement’ is the ‘Attitude-behavior Gap’ (f = 6) at the layer of actors, which refers to participation without achieving the activity’s goals, such as insufficient personal reflection. A notable subjective effect at the layer of actors is ‘Social Pressures and Anxiety’ (f = 4). For instance, while ‘Pokémon GO’ encouraged players to ‘play outside’ and facilitated ‘ice-breaking moments’, it also challenged their physical abilities and heightened social anxiety [109]. The effects at the relations layer are mainly related to social-cultural conflicts. One effect is ‘Unmatching Cultural Values and Communication Styles’ (f = 5). Notably, two cases highlight negative effects due to limited verbal cues: Taylor et al. found that the synchronous use of verbal and written communication can confuse users [155], and Bruža et al. noted users’ preference for familiar interaction methods such as rotating objects with both hands, similar to smartphone swipe interactions [181]. This indicates that focusing solely on verbal cues while ignoring non-verbal cues is insufficient for MSE, as detailed in Section 5.3. Another notable effect is ‘Non-compliant Social Norms and Ethics’ (f = 4). For example, the system ‘MagicART’ reported these issues when group members had different visiting expectations and styles [155].
- (2) XR-related effects. ‘Technological Bottleneck’ (f = 22) is the most frequently reported negative effect, highlighting the gap between technological expectations and actual implementation due to hardware and software limitations. Compared to VR, AR and MR more commonly face this issue, particularly in early-stage prototypes; VR encounters this less frequently. However, notably, ‘Physical Load’ (f = 10), typically caused by poor VR wearability, contrasts with MR systems using NUI and TUI, which rarely encounter such problems; see Section 5.2.3 Input Devices → Cues for details. VR also tends to cause ‘Sickness’ (f = 3). For example, Flynn’s VR study reported a support person experiencing a ‘tension headache’ due to the HMD and limited face padding of the Meta Quest Comfort Strap [184]. Among less frequent negative effects, ‘Ignorance of Privacy and Safety’ (f = 3) is notable, primarily concerning public MSE. Beyond addressing identity anonymity attacks in large VR social networks [112], attention should be given to privacy in public MSE. For example, Du noted that their design inadequately addresses privacy concerns, as participants highlighted a ‘lack of privacy’ during interactions with public displays, indicating the need for further research to ensure ‘personal space’ and mitigate interaction blindness in urban planning contexts [23]. Another public urban planning study echoed this point [118]. This point is detailed in Section 5.2.3 Avatars → Clustering. ‘Uncanny Valley Effect’ (f = 2) should not be overlooked. It now extends beyond avatars to real users; for instance, participants found face-mounted iPads in MR environments ‘weird’ or ‘creepy’ [125]. ‘Interference from Real World’ (f = 2) is also significant, typically occurring in VR systems, mainly due to sensory conflicts, as detailed in Section 5.2.3 Output Devices → Spatial Conditions.

5. Discussion and Conclusions

5.1. Limitations

This study has several limitations. To ensure reliable thematic analysis, structured codes and an inter-coder method were used to finalize the coding. However, coder subjectivity may still introduce bias. Additionally, to manage the volume of query results, explicit expressions of MSE and XR-related keywords were specifically used. Consequently, papers with implicit or domain-specific expressions, such as ‘digital twins’ or ‘natural user interfaces’, were excluded. This review also primarily focuses on English-language studies, potentially omitting relevant research published in other languages. Lastly, in analyzing RQ1.3, the sample size may have been insufficient, leading to statistically insignificant correlations for certain SF-mediated components, such as autonomy. While different aspects of autonomy, particularly its mediation by SFs, were observed in RQ1.1 and RQ2, the interaction with various types of immersion remains unclear. It is uncertain whether these interactions act as constraints or enhancements on specific immersion types. This limitation can be addressed with larger sample sizes in future research.

5.2. Major Findings and Discussions

5.2.1. Mechanism of XR-Mediated MSE

The distribution of each component and SFs were analyzed to answer RQ1. Firstly, diverse cognitive-behavioral performances in MSE were observed in answering RQ1.1. Some imbalances in the data distribution were identified:

Firstly, non-verbal communication modes in XR systems are still limited, yet several studies have demonstrated their potential to enhance social experiences. Tarnec et al.’s *VRInteract* system combines users’ actual non-verbal behaviors with artificial cues to improve social presence and attention during interactions [51]. By reproducing and augmenting non-verbal behaviors such as gestures, the system fosters a more fluid and intuitive communication environment. This approach directly taps into *ECT* [97], which posits that our physical actions and interactions with the environment play a critical role in enhancing cognitive processes. Similarly, Wu et al.’s *MR.Brick* game utilizes TUI and gestures, allowing participants, particularly children, to express themselves more naturally without relying solely on verbal communication [157]. This integration of physical interaction with digital space helps reduce cognitive load, thereby promoting smoother, more intuitive social exchanges. Common across these systems is the use of non-verbal cues to enhance engagement and collaboration by establishing a shared social space where users can interact more meaningfully. Both *VRInteract* [51] and *MR.Brick* [157] demonstrate that incorporating gestures and body movements—key elements in *ECT* [97]—facilitates better communication, especially in environments where verbal communication might be limited. Tsoupikova et al.’s multi-user VR system for post-stroke rehabilitation similarly relies on body motion tracking and avatars to recreate non-verbal interactions, which fosters social connection and engagement, even in remote settings [158]. These studies suggest that improving users’ ability to communicate non-verbally can enhance social presence, leading to more immersive social experiences in XR.

Additionally, although cases of mixed and asynchronous communication are currently limited, researchers have proof of its potential. Pre-recorded videos with voiceover [106,134] and geotagged information markers [118,147], as well as the real-time use of emojis and text [23,148,165,175], were widely utilized in mixed communication. These measures were usually designed for non-public engagement, focusing on improving interaction efficiency. The greater potential of asynchronous communication manifested in systems aimed at engaging a public suffused with unfamiliar relationships. Furthermore, the effectiveness of asynchronous communication lies in their ability to (1) Increase flexibility. These modes allow participation at any time, attracting a broader audience across different time zones, as seen in Koukopoulos’ AR-based library system [113] and Prandi et al.’s *ShareCities VR app* [119]. (2) Sustain thoughtful interaction. Users can contribute at their own pace, as evidenced by Koukopoulos’ system where readers leave detailed AR annotations, promoting

reflective engagement [113]. (3) Higher scalability. Higher scalability. These systems can support large, dispersed audiences, making them ideal for public or tourist-oriented applications such as Spiel et al.'s 'Way-finding Pillars' system [130]. The above examples confirm the potential and advantages of incorporating asynchronous communication modes into XR systems. However, regardless of the specific scenario, the following basic challenges are likely to arise when using asynchronous communication modes, and future research should take these into account: (1) Delayed feedback. The lack of real-time responses can reduce interaction immediacy, which may hinder active social engagement. (2) Integration complexity. Balancing asynchronous with synchronous modes can be difficult, particularly in managing the flow between live and pre-recorded interactions. (3) Maintaining engagement. The absence of instant feedback may lead to reduced user interest over time.

Moreover, while collaboration was a primary focus, the interplay between cooperation and competition, particularly 'coopetition', was often overlooked. Only five samples explored 'collaborative competition', primarily limited to the scenario of competition among cooperative subgroups in education [147,151] and *Player versus Environment (PVE)* [156,157,167] in games. An in-depth exploration of the coopetition mode is necessary.

Answering RQ1.2 revealed that the diverse combination of SF tags indicates a rich accumulation of technology-mediated strategies. For example, 'Input and output devices' integrate multi-interface systems to enable more users to access shared environments [167]. However, data distribution further suggests some potential gaps and opportunities, such as the emerging but underexplored role of 'AI agents' in facilitating interaction targets. Additionally, 'sensory feedback' showed a significant imbalance, with tactile feedback being underutilized and multisensory environments, including olfactory and gustatory modalities, remaining neglected (further refer to Section 5.2.3 Output Devices → Spatial Conditions). Data on 'absorption' indicated a strong inclination towards strategic immersion, though tactical and mixed immersion modes, crucial for enhancing memorable experiences and managing fatigue, were underutilized.

Addressing the identified data distribution imbalances is essential. It is not only to fill existing research gaps and expand new opportunities but also to promote diversity and equity in the development of XR-mediated MSE.

In answering RQ1.3, pairwise strongly correlated (SFs → components) relationships were explored. Some had positive impacts, such as (avatars → cues), while others had negative effects, such as (spatial conditions → absorption) and (topology → sensory feedback). This highlights that mediated relationships do not fully apply to all contexts and exhibit both positive and negative outcomes. While acknowledging the positive aspects, it is equally important to critically reflect on the accessibility of mediation strategies and their potential negative impacts. Due to the critical nature of these relationships, further elaboration is provided, as detailed in Section 5.2.3.

5.2.2. Effectiveness of XR-Mediated MSE

Addressing RQ2.1 reveals that the domain of XR-mediated MSE has evolved to a point where evaluation settings are reaching maturity. There is a discernible shift towards a more holistic approach, with a greater emphasis on MSE aspects rather than just the performance of XR technologies.

Quantitative methods are the mainstay of these evaluations, yet they are consistently paired with qualitative insights to achieve a well-rounded assessment. A variety of methods are utilized, balancing quantitative tools such as questionnaires and controlled experiments with qualitative approaches such as interviews and observations. All of these methods are essential for gathering detailed user feedback. The introduction of novel methods, such as VFT for exploratory VR activities and CBPR for collaborative initiatives, enriches the evaluation process by tailoring assessments to specific contexts. Nevertheless, the sparse application of physiological measurements is a notable drawback, given their potential to provide objective markers of social behavior—a dimension that warrants future attention. The preference for mixed methods and the development of scales such as the GEQ and SUS

signify an evolving framework for evaluation, yet challenges remain. These include the inherently subjective nature of MSE metrics and the pressing need for a unified approach to XR system metrics.

Looking ahead, the focus should be on refining these scales, enhancing the objectivity of MSE evaluations, and establishing a consensus on XR metrics. These efforts will bolster the methodological rigor and applicability of future research in the field. In summary, while the initial patterns of evaluating XR-mediated MSE have begun to solidify, there is a clear necessity to develop a standardized assessment toolbox. This will provide researchers with a reference knowledge base and expand the foundation of evaluative practices.

Addressing RQ2.2 involved examining the effectiveness of XR-mediated MSE from both social and technological perspectives. In addressing RQ2.2, the study evaluated the effectiveness of XR-mediated MSE from both social and technological perspectives. Socially, XR platforms have demonstrated the potential to enhance group engagement, stimulate social motivations, and strengthen team cohesion. However, challenges such as disengagement and the attitude-behavior gap were noted, revealing the need for equitable social structures and strategies to reduce social pressures. Technologically, XR-mediated MSE shows improved task efficacy, user experience, and technical accessibility, but ongoing limitations such as technological bottlenecks and physical discomfort from devices persist. Moving forward, overcoming these challenges will be essential for further advancing the field.

While XR-mediated MSE has made progress in enhancing both social interaction and technical performance, significant challenges remain. Future research should focus on improving system usability, minimizing physical discomfort, and ensuring seamless and intuitive interaction. Addressing these areas, along with privacy and ethical concerns, will help unlock the full potential of XR technologies for more effective social engagement in MSE environments.

To further deepen the answers of RQ1 and RQ2, Section 5.2.3 analyze significant case studies related to mediated relationships.

5.2.3. Significant Mediated Relationships

Output Devices → Spatial Conditions

Output devices play a crucial role in enhancing spatial conditions. '*Positive Enthusiasm and Scale of Engagement*' as reported positive results are mainly derived from studies on group-level or public engagement with AR/MR systems, as reported in representative cases [115,130,133,141,150,156,163]. Factors such as low device access requirements (especially mobile phones/tablets), easy navigation, and reality-based environments, enable MSE in a low-cost manner. Particularly, output devices such as tabletop projections [135,137,146,151,171], wall-based displays [23,141,161,170], 360-degree theaters [153,178], and floor-based forms [156] have been instrumental in supporting collocated MSE. Users can manipulate virtual objects in closer physical proximity, demonstrating the positive impact of output devices on spatial conditions and social interaction cues, thereby fostering active exploration. VR systems also have this potential by presenting an artificial world that fully occludes the physical environment, offering a more immersive audiovisual experience with meaningful content that links hybrid and remote users (see cases: [46,175,180,181,184]).

While audiovisual output interfaces have advanced, integrating multisensory feedback remains a significant challenge, largely due to current technological barriers in output devices. Tactile, olfactory, and gustatory feedback are underdeveloped, which limits the perception richness of social spaces. Tactile feedback, often limited to basic haptic devices such as joysticks, provides minimal sensations such as force or vibration, with few exceptions such as tactile social robots for close-range interaction [127] and whole-body floor vibration platforms [114]. Olfactory and gustatory feedback are even more neglected, despite some early attempts, such as Meenar and Kitson's use of scents (e.g., freshly cut

grass) in physical environments, which has shown potential in enhancing immersion and spatial exploration [178].

However, these sensory dimensions remain largely unexplored in mainstream XR applications, leaving implications to users' social experience: (1) Weakened emotional involvement. The absence of olfactory feedback can reduce emotional engagement and limit spatial exploration in MSE. In real-world interactions, scents play a critical role in conveying emotions and intentions. With olfactory cues, XR users could maintain focus, avoid boredom, and even activate their motor systems. For instance, in Meenar and Kitson's study, a user noted, "*I think it was very effective... the different smells and being able to see things in all directions was better*" [178]. (2) Disrupted embodiment in communication. Embodiment, or the sense that an avatar is the source of one's sensations, can suffer due to insufficient sensory feedback [99]. This sense of embodiment is typically achieved through synchronous visuotactile or visuoproprioceptive stimulation [194]. When users can only see and hear each other but cannot physically interact with the virtual environment, they may experience detachment from their avatars. This sensory gap can diminish the immersiveness of social interactions, as touch plays a crucial role in grounding users in virtual spaces.

While multisensory integration is critical, the output content at the same sensory modality level is also worth noting, particularly auditory conflicts arising from real-world noise interference. Some negative user feedback about VR output information conflicts in samples has been highlighted: '*The robot vacuum cleaner whizzing around the virtual living room*' [189]. These disruptions can affect attention and break presence and immersion.

Input Devices → Cues

Multi-interface XR systems have significantly advanced the accessibility of social cues while expanding engagement by offering user-preferred modes. The system 'PaKOMM' developed by Postert et al. [152] exemplified this by integrating an interactive touch table with VR/AR devices. This system allowed users with different roles to sequentially engage with various interfaces. They started at the touch table to discuss and outline initial plans. Then, they moved into VR/AR environments to further develop and iterate on these ideas. Finally, they returned to the tables to review the outcomes. This approach leverages the unique operational advantages of each interface, enhancing the utility of social cues. Similarly, the 'ShareVR' system by Gugenheimer et al. [144] synchronized VR and MR experiences, incorporating real-time floor projections with position tracking. The tracking systems transmitted MR users' cues to VR users, allowing non-HMD users to actively understand MR users' intentions. Another example is the 'XRPublicSpectator' system. It mapped the virtual cues input by VR performers onto a public screen, allowing non-HMD users to more comprehensively perceive the detailed cues of the VR world, such as object control, within the MR environment. This approach prevents the imbalance that can result in inefficient presenter-spectator communication and reduced spectator engagement. More relevant cases can be found in [125,165,167]. In summary, multi-interface XR systems connect users of different interfaces. They not only enhance the expression of social cues by providing user-preferred interfaces but also ensure the wider accessibility of engagement in environments with limited equipment.

NUI and TUI enhance social cue expression in MSE by making interactions more intuitive and natural. In MR environments, recognition systems link users' posture and physical object-related data with virtual objects. For instance, the 'SAR-Connect' system [164] allows stroke patients to interact with social robots and virtual elements simultaneously, incorporating hand-state-recognized finite-state machines for conversational and training convenience in rehabilitation. Additionally, Samayoa et al.'s system [123] uses holographic displays to process 3D object data (e.g., color and position) for remote users, enhancing their ability to convey behavioral intentions with fewer barriers. Similar enhancements are observed in other MR systems [161,170]. These systems use tracking systems as input media, enabling users to interact without physically touching control devices, thus shifting attention from controlling devices to socializing and reducing bodily load. The integration

of tangible components further supports SI. Examples include a tangible construction set featuring boards, cards, and clickers [157], NFC-tagged plastic bouquets [136], and touch tables [129,152]. These components create a direct link between actions and results, reducing the cognitive and physical cost of understanding interactions, and thereby enhancing social engagement. The music co-creative system '*BilliArT*' [146] supplies billiard sticks for strokes that trigger jazz improvisations. It utilizes a straightforward interactive device (the billiard) to establish sound-action associations, which are integrated into a multimodal audio-visual feedback system. This setup stimulates novel and engaging physical expressions and mutual communications. For example, it provides aesthetic involvement, encouraging users to appreciate the installation as a whole rather than focusing separately on the game, music, or visuals. TUI facilitates direct interaction with the digital world by augmenting physical objects, surfaces, and spaces, enhancing users' natural dexterity by materializing digital data in physical space. Their integration into XR systems significantly boosts users' ability to sense and manipulate these systems, creating richer social experiences.

ECT [97] strongly supports the advantages of NUI and TUI. It posits that cognitive processing is shaped by the entire body, underpinned by perceptual and motor systems and bodily-environmental interaction [195]. Augmented bodily states and modality-specific systems for perception and action supplied by NUI and TUI reinforce information processing, demonstrating that embodiment contributes to the positive construction of mental phenomena [97]. In the future, XR-mediated MSE systems can further incorporate these two interfaces, especially for users with mobility impairments, such as the stroke patients in the '*SAR-Connect*' system [164].

Avatars → Cues

In XR systems without avatars, users primarily rely on verbal communication, object control, and posture to express cues. However, in VR systems with avatars, these avatars enable a broader range of cues, particularly in full-body mode, which enhances comprehensive cue expression in VR environments [132,162,179,184,188]. This illustrates that the use of avatars, especially full-body avatars typically viewed from a third-person perspective [132,154,162,179], integrates richer cues. Users reported feeling genuinely immersed in the environment with others via avatars, as noted by [143]: '*It felt like you were really in the world with the other people. The other avatars also talked to you normally, so it felt like you were really in a room with other people*', and '*You did look at the avatar when you were talking, as you normally do in a conversation.*' Their positive role in interpersonal communication has also been confirmed by numerous VR studies [132,134,143,177,180].

Regarding avatar visual types, two main settings are generally featured: the display of the hands [111,134,155,177] and upper limbs [158], as well as full bodies typically viewed from a third-person perspective [132,154,162,179]. These visual representations, more stylized and animated compared to physical bodies, help users better adapt to cue usage, which is highly relevant to coopetition (see Section 5.2.3 Avatars → Coopetition). Additionally, avatars occasionally include non-human-like figures such as animal characters [134,148,177], as well as cursors [106], expanding personal identity and adding a playful element to SI, thereby strengthening social presence. From the *Hyper-personal Model* perspective [196], intimacy in XR systems is closely related to emotional fulfillment, the need to be touched, and interpersonal communication. XR-mediated intimacy introduces new forms of physical proximity and interaction through multimodal communication channels and avatar-mediated behaviors, generating positive user experiences and facilitating interpersonal relationships without the need for face-to-face interaction. Avatars anonymize or hide real-world identities, making it easier for users to feel closer to others and express cues. However, no significant differences were found in situations with asymmetric use of avatars (e.g., one user with an avatar, another without). Future research should address this aspect.

Avatars → Coopetition

In VR, partial-body avatar modes are more often observed by pairs of collaborative contexts, as exemplified by cases focusing on half-body avatars [46,51,148,181,189] or hand-centric avatars [111,134,144,177]. This mode simplifies visual processing, reduces computational and rendering complexity, concentrates cue expression, minimizes unnecessary body movements, and provides a more comfortable and efficient collaborative environment. This form is particularly advantageous in complex object distribution and spatial environments. For instance, Cassidy et al. [111] describe a VR prototype that helps archaeologists and other stakeholders explore and analyze archaeological data more immersively. Their collaborative environment required various complex interactions (e.g., using a flashlight tool to reveal original rock art), where semi-transparent hand and head presence avoided visual obstruction. Similarly, Shah et al. [46] considered peer collaboration needs when configuring avatars for players. They focused on upper-body movements, such as cervical, shoulder, elbow, and wrist movements. This led to the creation of half-body avatars to facilitate visual recognition and collaboration.

Full-body avatar modes are more often observed as pairs of neutral contexts where there is no significant coopetition, as shown in cases [119,132,155,179,184,188]. This mode usually fits large virtual scenes, offering greater freedom of choice and identity transformation, promoting social dialogues and narrative tasks. Each user could select their avatar, observable from a first-person perspective, as one would in the real world. When viewing each other in the VR, users could see entire bodies that could verbally talk, wave, and move toward one another in real time. Full-body avatars supported a higher perceived sense of closeness or social presence and awareness of each other's avatars in the environment. For example, in Flynn's study, user behaviors demonstrated this interaction: '*User A observed an instant shift in his body language. He started singing, waving his hands, and dancing with User B. They both sang together, smiling and laughing throughout the song. In the cinema room, they watched an 'Only Fools and Horses' video, and User C exclaimed, 'Delboy!' They laughed together while watching the comedy video.*' This type also occasionally applies to competition cases [114,167], mainly to provide contextual coherence between player and environment. For example, in Jung et al.'s case, participants were displayed as medieval knight avatars. They used hand-controlled 'aiming wheels' to manipulate a small-scale cannon and destroy their opponent's wall. This took place in a medieval tavern-like environment with lighting and ambient sound effects [114]. Personal traits could impact the perceptual response in multi-user VR experiences, with full-body avatars enabling better interaction within the contextual setting. In summary, half-body avatars are more often observed by pairs of collaborative tasks, enhancing task efficiency, while full-body avatars are better suited for contexts without strong coopetition needs, fostering a greater sense of affinity.

Avatars → Clustering

Personalized avatars enhance social clustering and relational dynamics by integrating various social mechanisms, primarily demonstrated in some VR scenarios. During the '*online IEEE VR Conference 2021*', Moreira et al. employed a '*Scientific Speed Dating*' session, encouraging users to manipulate their avatars to engage in fast-paced, informal, and energized conversations to establish trust and professional bonds [175]. Similarly, in the '*Icebreaker*' section of Gruber et al.'s study, personal avatars facilitated short dialogues that enabled participants to share personal stories and cultural insights, promoting interpersonal understanding and connection. This received positive qualitative feedback, such as: '*I was able to learn something about the other student's views, life, and culture*' [143]. These instances illustrate how avatars can bypass traditional, slow processes of relationship building, offering a platform for more dynamic and immediate connections. As Moreira reviewed [175], various conferences employed different media choices (such as Twitch, Hubs, and Slack) and assessed their appropriateness during the events. Their study indicated that for participants who experienced avatar-based VR platforms, perceived social presence and the desire for communication were highest due to the integration of

avatars. This contrasts with traditional video conferencing, which lacks the immersive, expressive, and personalized elements that avatars provide. Ultimately, this demonstrates that avatars contribute uniquely and significantly to social engagement.

However, the use of avatars has proven to be a double-edged sword in terms of social clustering within VR environments. While they can significantly enhance SI and network vitality, their misuse can lead to negative experiences. Maloney et al. highlighted how spontaneous, interest-based social clustering around avatars could also foster behaviors contrary to societal norms, including harassment and bullying. For example, an older teenage boy on VRchat stated, '*He deserves to die from the coronavirus by just looking like that [reference to his avatar]. That avatar is lame as fk, who the fk has a 1D avatar, haha.*' This was evident in interactions where participants commented negatively on each other's avatars or discussed inappropriate topics, leading to discomfort and adverse social dynamics.

Avatars can enhance social and spatial presence but may also have technical and application flaws. High costs, limited creativity, and oversight of deep information hinder innovation and integration into the industry value chain. Structured governance and ethical oversight in avatar-mediated interactions are essential to prevent discomfort and adverse social dynamics.

Activity Range → Spatial Conditions

Most studies utilized location-based services to integrate real-time GIS information and physical environments into XR. This static spatial presentation requires users to exert more physical effort for exploration. This can be challenging for users with mobility impairments, as it increases physical load and limits the diversity and capacity of information presentation.

Sample scenarios often involved movement across large communities [115,118,124,152,163], tourist hotspots [117,130], and globally accessible sites via mobile apps such as '*Pokémon GO*' [169]. While these locations allowed extensive physical freedom, they reduced efficiency and accessibility for those with mobility impairments. Innovative solutions have emerged to address these constraints. For example, Postert et al.'s '*PaKOMM*' system integrated real-time geographic data from different locations to facilitate shared experiences between remote users [152]. Similarly, the '*CommunitAR*' platform allowed residents to interact with urban features and share spatial data with decision-makers remotely, enhancing community engagement in urban planning [118]. An innovative study by Yavo-Ayalon et al. situated co-located participants in a traveling bus equipped with HMDs to immerse them in a climate change scenario [124]. This approach provided a virtual representation of real-world geographic information, enabling trans-spatial exploration. Additionally, pre-fabricated off-site scenarios, such as Meenar et al.'s holographic simulations for remotely visualizing landmarks [178] and Puggioni et al.'s '*ScoolAR*' platform for virtual tours [165], showcased the potential of transcending physical constraints. Both approaches combined motion and stillness, providing broader opportunities for social engagement.

All in all, static design methods within fixed contexts may overlook the benefits of dynamic approaches. Integrating static and dynamic information presentation methods can create novel experiences, allowing comprehensive spatial exploration and catering to a wider range of users.

Interaction Targets → Affects

Human-only interactions often display sporadic or intense emotional patterns. However, human+AI interactions exhibit more consistently intense emotional patterns [107, 153,167,178,190], reflecting changes in social perceptions and norms when AI is involved, as per CASA theory [89]. AI typically induces intense emotional arousal through novelty and proactive engagement. For instance, Gugenheimer et al.'s system '*BeMyLight*' [167] featured AI zombies that not only raised their claws menacingly but also dodged attacks, heightening tension and engagement among players. Conversely, pure human interactions, such as in VR E-learning environments [111,116,140], often involve slower emotional pro-

cesses. Therefore, interaction targets and affect are strongly correlated. Further speaking, the deployment of AI as additional ‘interaction targets’ significantly enhances social ‘affect’, particularly in shaping empathy and embodiment.

Empathic AI enhances users’ emotional and behavioral investment in interactive systems by expressing complex emotions such as hopelessness [153], fear and stress [167], and a need for comfort [127]. For instance, Pedram et al. [153] used AI to simulate stranded survivors in a mine rescue scenario. This effectively captured trainers’ attention and accelerated cognitive processes through realistic depictions of hazards such as fire and toxic gases. Similarly, Papadimitriou et al. [154] developed the ‘HTML Escape Game’, where players helped an AI ‘hero’ escape from a castle. This educational entertainment project engaged players with novel-style dialogues and HTML-related puzzles, enhancing their problem-solving motivation. Bird et al. [107] fostered increased cognitive and emotional engagement by integrating unexpected events in their museum game, where players assisted AI ghost dinosaurs in returning to ‘ghost land’.

Social robots can be perceived as the physical representation of AI in XR systems. Compared to screen-based or virtual AI, they more effectively foster healing in mental health interventions, generate positive emotions, and enhance social connections [197]. For instance, in Feng et al.’s study [127], their robot served as a communicative intermediary, supplemented by virtual content displaying dynamic AI sheep. Disguised as a lamb and equipped with furry textile and soft stuffing, the robot engaged closely with patients by exhibiting joyful gestures when touched (e.g., happy gestures expressed using its head, neck, legs, and tail), provoking playful experiences of sensation, relaxation, and reminiscence, thus enabling deep emotional intervention. Overall, compared to human-mode MSE, AI agents typically enhance openness, agreeableness, and self-disclosure rapidly. Future research can focus on the role of physical AI or social robots in XR environments and their impact on user emotional stimulation and guidance.

Interaction Targets → Coopetition

AI often facilitates collaboration more effectively than pure human interactions, which may involve lower willingness and goal orientation among unfamiliar participants. Meenar and Kitson [178] demonstrated that AI-simulated urban sounds subtly influenced the collaborative mood among urban planners. Other implementations highlight AI’s role as ‘narrative guiders’ [117,159], ‘reciprocal companions’ [127,164], and even ‘extras’ [178]; through these new role allocations, the efficiency of collaborative tasks has been enhanced. In ‘Lands of Fog’, Sayis et al. [156] used AI to transform players’ appearances into rhythms, engaging autistic players with peers in gathering AI fireflies, eliciting reflective responses and increased thrill-seeking. However, excessive AI usage can lead to distractions and dilute the collaborative atmosphere, as noted by Meenar and Kitson [178].

All samples placed AI in the role of conversation and emotion agents, facilitating simple forms of collaboration without emphasizing their practical deep functionality or considering human–AI collaboration. Notably, the integration of concepts such as AI-generated content (AIGC), ChatGPT, and Large Language Models (LLM) into the XR environment was not explored, with AI primarily acting as a mediator of personal and group interactions. Recent advancements in AIGC bring life to the metaverse by dynamically creating content for non-playable characters. This includes head generation, full-body non-verbal expressive features, clothed NPC creation, dance animation, and animal avatar generation [198]. These functions were not reflected in the samples, which may underestimate the value of AI agents in MSE.

Spatial Conditions → Absorption

Objective spatial conditions influence the selection of tasks that generate different types of absorption. Observations show that absorption is often correlated with specific spatial conditions. Collocated settings often involve strategic and narrative immersion [107,116,154,157,173,177], such as in Doukianou et al.’s integration of strategic content

with narrative interactive infographics in AR business presentations, which enhances spatial cohesion [150]. Remote and distributed conditions involve more narrative and tactical immersion, increasing engagement among users in these settings. Integrating tactical immersion effectively combats the potential monotony of a single immersive experience, particularly in enhancing narrative immersion. This approach was exemplified by the modular deconstruction and integration of linear narrative contents into the walking exploration of physical spaces. For example, AR interactions at Badaling National Forest Park linked thematic stories to the environment [117], guiding users to explore and solve puzzles, and demonstrating how task-induced absorption affects remote and hybrid conditions. Another notable implementation was seen in Lee's mobile gaming application '*Neighborhood History Walk*', which guided community members through their city, enhancing engagement with historical photographs and neighborhood stories [163]. Their investigation showed that narrative-walking integration not only brought fun and excitement but also enhanced the recreational experience and made tourism more memorable. Therefore, when implementing an absorption strategy, spatial conditions must not be overlooked. They should be considered to avoid making the technology-mediated strategy inaccessible. Additionally, leveraging the advantages of spatial conditions can lead to more effective absorption.

Furthermore, a critical aspect that requires careful consideration is the scale of absorption-affected mental and physical states. For instance, one participant in VR space exploration remarked about the highly realistic spatial immersion, '*It's not necessarily positive for everybody. And even that sense of floating, for some people, that's great, for other people it's disconcerting*' [142]. Bekele's study also highlighted user feedback suggesting a reduction in text and video content. It noted that lengthy content might fail to capture visitors' attention [173]. This reflects that strategic immersion should also avoid information overload.

Topology → Sensory Feedback

Topology limits the implementation of multisensory outputs, much like spatial conditions potentially constrain absorption (see Section 5.2.3 Spatial Conditions → Absorption). Auditory and visual channels can adapt to various topologies, while haptic feedback is limited to point-to-point topology due to high development costs, restricting its application in large social settings. For instance, emerging tactile cues such as whole-body feedback from floor vibrations enhance perceptual responsiveness between dyadic users but are not yet feasible for large social scenarios [114]. In contrast, olfactory channels are often observed to be paired with large, multi-user settings due to lower costs [178]. Thus, topology constrains the achievable sensory feedback, and the desired MSE structure limits technological deployment, indicating a reciprocal influence. As Sørensen notes in *Mediation Theory*, '*Which has real being, also, has the possibility to determine and/or constrain technological mediation—such as what is physical and/or technically possible, user context, cognitive structures (e.g., the interpretants), as well as sociocultural conditions (we know that these determinants/constraints often are intertwined)*' [199]. Sometimes, XR lacks the ability to mediate MSE effectively due to the predefined topology type. Notably, as the sample size within the framework increases, similar limitations in the enhanced relationships may also arise, which future research should address.

5.3. Future Research

Future research could focus on two main aspects: the mechanisms and the effectiveness of MSE. Based on the prioritization of findings, the following issues are particularly urgent in terms of the mechanisms:

- **Non-verbal communication.** Future studies should focus on refining these non-verbal communication systems to further align with the principles of ECT [97]. Specifically, exploring how different types of gestures, facial expressions, and postures can be systematically integrated into XR platforms to improve interaction quality will be crucial. Additionally, understanding how diverse user groups, particularly those with

- communication impairments or physical disabilities, engage with these non-verbal features will be key to expanding the inclusivity of XR experiences;
- **Mixed and asynchronous communication.** Key considerations should include managing delayed feedback, as the lack of real-time interaction can hinder social engagement. Additionally, integrating asynchronous and synchronous modes seamlessly remains challenging, especially in balancing live and pre-recorded content. Finally, maintaining long-term user interest is essential, as the absence of immediate feedback may lead to disengagement over time;
 - **Sensory feedback.** Although audiovisual interfaces have advanced, technological limitations still hinder the implementation of multisensory feedback, particularly in tactile, olfactory, and gustatory areas. This restriction weakens emotional engagement and immersion. Furthermore, when the same sensory modality is simultaneously triggered by both virtual and real-world inputs, it can lead to sensory disruption and confusion for users, breaking their sense of presence in the virtual environment. Meanwhile, due to the counteracting force of social topology on MSE activities, tactile interactions at the group level and in public MSE may be difficult to achieve. When considering technology mediation, social topology should be reasonably deployed in activities to avoid technological inaccessibility.

In addition to the aforementioned critical issues, future research should also focus on the following aspects:

- **Coopetition modes.** Collaborative competition and competitive collaboration have not been fully applied in coopetition, despite their significant importance in expanding group dynamics. It would be beneficial to compare the differences and similarities of various forms of coopetition further and their applicability to specific scenarios;
- **AI agents.** AI agents have a significant role in enhancing intensive affect mode and user empathy, yet this has not been emphasized. Designing vivid AI with customization and context awareness can deepen emotional engagement. The mediation of emotional modes still has much to explore, such as expanding AI's role in emotional communication, regulation, analysis, expression, and deeper reflection. Currently, AI agents in MSE primarily focus on emotion mediation and collaboration guidance, with their broader functionality not fully realized;
- **Immersion.** Current research emphasizes guiding users into narrative and strategic immersion states, while the potential of tactical immersion has not been sufficiently highlighted. Additionally, excessive immersion may cause cognitive or physical overload. Researchers could further define a rule library for immersive task design to avoid negative effects;
- **Multi-interface XR systems.** Multi-interface XR systems effectively mediate the engagement of asymmetric groups, enhance user preferences, bridge the interruption of cue expression in asymmetric users, and align task execution, warranting further attention. More attempts to introduce MR systems in conjunction with VR devices can further enhance broader multi-person engagement under limited VR equipment conditions.
- **NUI and TUI.** NUI and TUI not only enhance system manipulation perception, reduce users' physical and cognitive burdens, and support the social engagement of minority groups but also allow users to express more interesting cues, increasing the sense of enjoyment and promoting memorable experiences. Future work could further incorporate these interfaces into XR systems to create more embodied experiences;
- **Dynamic spatial design.** Static design methods within fixed contexts may overlook the benefits of dynamic approaches. Geographic location data leads to static information presentation, which fails to meet diverse user needs. Dynamic spatial design methods should be further expanded in XR spatial experiences to offer broader scene participation. An example can be seen in InnoVision Inc's recent development of the *MOON VR HOME app* [200]. With just a few clicks, users can transform their virtual home into a stunning landscape, a sci-fi game scene, or an interior masterpiece [200];

- **Avatars' symmetry.** As virtual extensions and reshaping of physical identities, avatars can enhance users' emotional satisfaction and social desires. Avatars enable users to focus more on expressing and perceiving others' cues and enhance the sense of presence, social presence, and spatial presence. No significant differences were found in situations with asymmetric use of avatars (e.g., one user with an avatar, another without), which requires further attention in future studies. Further exploration of visual symmetry differences and mediation strategies in specific task modes and contexts, along with defining rule sets, can maximize avatars' expressive value as cue agents in SI;
- **Avatars' visual types.** Half-body avatars are more often observed as pairs of collaborative, interactive content. They reduce computational and rendering complexity, concentrate cue expression, minimize unnecessary body movements, and provide a more comfortable and efficient collaborative environment. Full-body avatars are more neutral contexts. There is no significant competition. This supports higher presence and a perceived sense of closeness or social presence. It also increases awareness of each other's avatars in the environment. These conclusions are preliminary and require a further systematic guidance framework;

For effectiveness aspects, there are three major issues that may need attention:

- **Standardized evaluation system.** The lack of consensus in XR-mediated MSE evaluation frameworks hampers the reproducibility and comparability of findings. Future studies should address the following key areas: (1) Standardization of Core Metrics and Taxonomies: Current metrics should be further standardized and classified using taxonomies tailored to specific XR types, MSE scales, and domains. This involves applying feature engineering to refine evaluation toolboxes, ensuring greater granularity and selectivity. Building a knowledge graph of these metrics will help researchers map their work to existing knowledge and improve reproducibility. Effect size evaluation and further meta-analysis should support this process. (2) Comprehensive Measurement Integration: Current studies often fail to integrate both XR-related and MSE-related metrics comprehensively, limiting the potential to uncover blind spots in evaluation systems. Future research should focus on more holistic measurement approaches. (3) Mixed-Methods Approach: Physiological measurements can enhance objectivity and complement subjective assessments (e.g., questionnaires, interviews, self-reports). However, their application remains limited. Future evaluations should combine quantitative (e.g., physiological data) and qualitative (e.g., user feedback) methods, which can reduce bias and improve comparability. (4) Benchmark Datasets: Developing shared datasets across studies allows for benchmarking and replication. These datasets will facilitate cross-system comparisons and ensure cumulative research builds on existing findings. (5) Interdisciplinary Collaboration: Researchers from fields such as HCI, psychology, behavioral science, and computer science should collaborate to co-create standardized evaluation guidelines. Such interdisciplinary input ensures robustness and broader applicability of the evaluation system. (6) Evaluation Communities and Iterative Validation: Given the contextual complexity of XR-mediated MSE, regular validation of evaluation systems across different platforms and user groups is essential. Establishing evaluation communities and continuously updating knowledge repositories will ensure knowledge dissemination and iterative improvement;
- **Ethical concerns of social identity.** Ethical issues arise across XR platforms with varying levels of technological maturity and require targeted solutions. In highly developed VR social networking platforms, immersive virtual environments and full-body tracked avatars enhance social presence but also increase risks such as verbal harassment and bullying. These behaviors can negatively affect users, particularly minors, who are more susceptible to identity distortion and 'false memories' due to their developmental stage [112,201]. Current mitigation efforts focus on cybersecurity education, behavior management (e.g., account suspensions), and ethical guidance from guardians [112]. However, they often overlook the presence of negative identities

related to avatars. One alternative is to replace human-like avatars with non-human, abstract social cues, such as biosensory data that anonymize users by representing physiological inputs as simple visual elements, i.e., biosensory cues [202]. This can help reduce biases linked to race, gender, and class while safeguarding user privacy and promoting a healthier social dynamic within VR spaces [203]. In some AR/MR applications, which are more task-oriented and at an earlier stage of development, insufficient attention to ethical issues in interface design can lead to power imbalances in public collaborative environments. Stakeholders, such as planners or designers, may resist equitable power distribution, creating a ‘tree topology’ that restricts information flow and limits engagement [118]. Interface designs should aim to provide both shared and personal spaces, balancing power dynamics and reducing social conflicts that arise from identity exposure [23]. In summary, future research should explore the extent to which personal identity is represented across different interfaces and task types in XR systems. Uncontrolled display of identity information may further exacerbate social and ethical issues;

- **Privacy breaches of technical platforms.** Privacy breaches in XR platforms vary depending on device type, necessitating tailored data protection strategies. For HMD-based VR, end-to-end encryption is critical, as these systems often track sensitive biometric data (e.g., gaze and body movements). AI-driven privacy measures can help anonymize this data, reducing leakage risks while maintaining immersion. An end-to-end communication approach—where data is transmitted directly without server storage—minimizes third-party risks but increases the demand for device-side processing [204]. In mobile-based AR and XR systems (e.g., prototypes) involving sensor integration, device-level biometric safeguards such as facial recognition and fingerprint authentication, combined with user-controlled settings, are essential for protecting personal information. Simpler XR systems should prioritize transparent consent and granular data-sharing options, allowing users to manage their exposure in public or shared spaces. In short, privacy protection in XR requires device-specific approaches. Future research should further explore data protection measures for input devices to avoid potential ethical issues.

Meanwhile, the following issues should not be ignored:

- **Long-term engagement.** XR-mediated MSE should not only focus on short-term engagement but also on more sustainable and spontaneous MSE. Therefore, while evaluating the effects of technology mediation, long-term studies should also be considered;
- **Expansion and iteration of emerging methods.** Recent research has introduced emerging, XR-specific methods such as VFT and CBPR. However, their unique advantages and effectiveness remain underexplored. For example, CBPR methods should be further leveraged to manage stakeholder power dynamics while safeguarding participant privacy, thereby ensuring methodological integrity. These approaches require further expansion and iterative refinement to fully realize their potential in XR-mediated contexts;
- **Physical and cognitive load.** Details such as interference from the physical world, motion sickness, and physical load in VR can affect the accuracy and efficiency of effectiveness verification and should be addressed. A user-friendly technical guide might be necessary during experiments;
- **Cultural and social differences.** Cultural values, social personalities, expectations, and evaluations of MSE groups may lead to social misunderstandings, lack of motivation, and tension, affecting interaction effectiveness. Members should be more strategically organized, and attention should be paid to the engagement behavior of socially anxious or fearful users to enhance inclusiveness in group interactions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/systems12100396/s1>, Supplementary File S1. PRISMA Flow Diagram; Supplementary File S2. Sample Screening and Tagging Processes; Supplementary File S3. Detail Pairing Analysis of SFs and Components.

Author Contributions: Y.W. and R.X. were responsible for the conceptual and analysis framework design. Y.W. and R.X. and D.G. designed the data collection method. H.Z. and X.W. assisted with the data collection. Y.W., H.Z. and X.W. worked on data and result analysis. Y.W. was responsible for the initial writing of the manuscript. R.X. and D.G. directed and managed the entire study and was responsible for revising the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Ruowei Xiao's personal research start-up funds from the Southern University of Science and Technology, funding No. 65/Y01656112.

Institutional Review Board Statement: The study was approved by the Institutional Review Board (IRB) of Southern University of Science and Technology, Decision No. 20230122, Approval No. 2023PES122. The review was not registered. A protocol was not prepared, and there were no amendments to the information provided at registration or in the protocol.

Data Availability Statement: Data can be made available upon written request to the corresponding author.

Conflicts of Interest: There are no conflicts of interest among the authors of this manuscript.

Notes

- 1 <https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=terms-cramrs-v> (accessed on 1 September 2024). This study adhered to the IBM SPSS manual's effect size standards for Cramér's V, excluding variable combinations with $p \geq 0.05$ due to weak interpretative value. Specifically, results were categorized as follows: $ES \leq 0.2$ indicated a weak correlation, despite statistical significance, the fields were only weakly correlated; $0.2 < ES \leq 0.6$ indicated a moderate correlation; and $ES > 0.6$ indicated a strong correlation.
- 2 <https://snap.stanford.edu/class/cs224w-readings/brandes01centrality.pdf> (accessed on 1 September 2024).

References

1. Stephenson, N. *Snow Crash*; Spectra: Hong Kong, China, 2003; ISBN 0-553-08853-X.
2. Mystakidis, S. Metaverse. *Encyclopedia* **2022**, *2*, 486–497. [[CrossRef](#)]
3. Zhang, G.; Cao, J.; Liu, D.; Qi, J. Popularity of the Metaverse: Embodied Social Presence Theory Perspective. *Front. Psychol.* **2022**, *13*, 997751. [[CrossRef](#)]
4. Dolata, M.; Schwabe, G. What Is the Metaverse and Who Seeks to Define It? Mapping the Site of Social Construction. *J. Inf. Technol.* **2023**, *38*, 239–266. [[CrossRef](#)]
5. Founder's Letter, 2021. *Meta*, 28 October 2021.
6. Statista Metaverse—Worldwide | Statista Market Forecast. Available online: <https://www.statista.com/outlook/amo/metaverse/worldwide> (accessed on 4 January 2024).
7. Hennig-Thurau, T.; Aliman, D.N.; Herting, A.M.; Cziehso, G.P.; Linder, M.; Kübler, R.V. Social Interactions in the Metaverse: Framework, Initial Evidence, and Research Roadmap. *J. Acad. Mark. Sci.* **2022**, *51*, 889–913. [[CrossRef](#)]
8. Oh, C.S.; Bailenson, J.N.; Welch, G.F. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Front. Robot. AI* **2018**, *5*, 409295. [[CrossRef](#)] [[PubMed](#)]
9. Riva, G.; Galimberti, C. Computer-Mediated Communication: Identity and Social Interaction in an Electronic Environment. *Genet. Soc. Gen. Psychol. Monogr.* **1998**, *124*, 434–464.
10. D'Ausilio, A.; Novembre, G.; Fadiga, L.; Keller, P.E. What Can Music Tell Us about Social Interaction? *Trends Cogn. Sci.* **2015**, *19*, 111–114. [[CrossRef](#)]
11. Chetouani, M.; Delaherche, E.; Dumas, G.; Cohen, D. Interpersonal Synchrony: From Social Perception to Social Interaction. In *Social Signal Processing*; Burgoon, J.K., Magnenat-Thalmann, N., Pantic, M., Vinciarelli, A., Eds.; Cambridge University Press: Cambridge, UK, 2017; pp. 202–212.
12. De Jaegher, H.; Di Paolo, E.; Gallagher, S. Can Social Interaction Constitute Social Cognition? *Trends Cogn. Sci.* **2010**, *14*, 441–447. [[CrossRef](#)]
13. Berger, P.; Luckmann, T. The Social Construction of Reality. In *Social Theory Re-Wired*; Routledge: London, UK, 2016; pp. 110–122.
14. Liu, S. Social Spaces: From Georg Simmel to Erving Goffman. *J. Chin. Sociol.* **2024**, *11*, 13. [[CrossRef](#)]
15. Johnston, K.A. Toward a Theory of Social Engagement. In *The Handbook of Communication Engagement*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2018; pp. 17–32. ISBN 978-1-119-16760-0.

16. Xiao, R.; Wu, Z.; Buruk, O.T.; Hamari, J. Enhance User Engagement Using Gamified Ineternet of Things. In Proceedings of the Hawaii International Conference on System Sciences, Maui, HI, USA, 5–8 January 2021.
17. Homans, G.C. Social Behavior as Exchange. *Am. J. Sociol.* **1958**, *63*, 597–606. [[CrossRef](#)]
18. Bowlby, J. *Attachment and Loss*; Random House: New York, NY, USA, 1969.
19. Collins, R. *Interaction Ritual Chains*; Princeton University Press: Princeton, NJ, USA, 2004.
20. Albert, M. Luhmann and Systems Theory. In *Oxford Research Encyclopedia of Politics*; Oxford University Press: Oxford, UK, 2016.
21. Turner, J.H.; Maryanski, A. *Functionalism*; Benjamin/Cummings Publishing Company: Menlo Park, CA, USA, 1979.
22. Newman, B.M.; Newman, P.R. Part III—Introduction. In *Theories of Adolescent Development*; Newman, B.M., Newman, P.R., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 245–250. ISBN 978-0-12-815450-2.
23. Du, G.; Degbela, A.; Kray, C. User-Generated Gestures for Voting and Commenting on Immersive Displays in Urban Planning. *Multimodal Technol. Interact.* **2019**, *3*, 31. [[CrossRef](#)]
24. Xiao, R.; Wu, Z.; Hamari, J. Internet-of-Gamification: A Review of Literature on IoT-Enabled Gamification for User Engagement. *Int. J. Hum.-Comput. Interact.* **2022**, *38*, 1113–1137. [[CrossRef](#)]
25. Helliwell, J.; Putnam, R. The Social Context of Well-Being. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2004**, *359*, 1435–1446. [[CrossRef](#)]
26. Jackson, M. Things Hidden Since the Foundation of the World. In *Life Within Limits: Well-Being in a World of Want*; Duke University Press: Durham, NC, USA, 2011; pp. 63–76. ISBN 978-0-8223-4892-4.
27. Bales, R.F.; Strodtbeck, F.L. Phases in Group Problem-Solving. *J. Abnorm. Soc. Psychol.* **1951**, *46*, 485–495. [[CrossRef](#)]
28. Krause, M.S. Use of Social Situations for Research Purposes. *Am. Psychol.* **1970**, *25*, 748–753. [[CrossRef](#)]
29. Moos, R.H. Conceptualizations of Human Environments. *Am. Psychol.* **1973**, *28*, 652–665. [[CrossRef](#)]
30. Price, R.H.; Blashfield, R.K. Explorations in the Taxonomy of Behavior Settings. *Am. J. Community Psychol.* **1975**, *3*, 335–351. [[CrossRef](#)]
31. Forgas, J.P. 1—The Perception of Social Episodes: Categorical and Dimensional Representations in Two Different Social Milieus. *J. Personal. Soc. Psychol.* **1976**, *34*, 199–209. [[CrossRef](#)]
32. King, G.A.; Sorrentino, R.M. Psychological Dimensions of Goal-Oriented Interpersonal Situations. *J. Pers. Soc. Psychol.* **1983**, *44*, 140–162. [[CrossRef](#)]
33. Kreijns, K.; Kirschner, P.; Jochems, W.; Buuren, H. Determining Sociability, Social Space, and Social Presence in (A)Synchronous Collaborative Groups. *Cyberpsychol. Behav. Impact Internet Multimed. Virtual Real. Behav. Soc.* **2004**, *7*, 155–172. [[CrossRef](#)]
34. Vrieling-Treunter, E.; Henderikx, M.; Nadolski, R.; Kreijns, K. Facilitating Peer Interaction Regulation in Online Settings: The Role of Social Presence, Social Space and Sociability. *Front. Psychol.* **2022**, *13*, 793798. [[CrossRef](#)] [[PubMed](#)]
35. Wu, H.; Liu, X.; Hagan, C.C.; Mobbs, D. Mentalizing during Social InterAction: A Four Component Model. *Cortex* **2020**, *126*, 242–252. [[CrossRef](#)] [[PubMed](#)]
36. Meijerink-Bosman, M.; Back, M.; Geukes, K.; Leenders, R.; Mulder, J. Discovering Trends of Social Interaction Behavior over Time: An Introduction to Relational Event Modeling. *Behav. Res. Methods* **2023**, *55*, 997–1023. [[CrossRef](#)] [[PubMed](#)]
37. Yang, Y.; Read, S.J.; Miller, L.C. A Taxonomy of Situations from Chinese Idioms. *J. Res. Personal.* **2006**, *40*, 750–778. [[CrossRef](#)]
38. Pervin, L.A. Definitions, Measurements, and Classifications of Stimuli, Situations, and Environments. *Hum. Ecol.* **1978**, *6*, 71–105. [[CrossRef](#)]
39. Hoppler, S.S.; Segerer, R.; Nikitin, J. The Six Components of Social Interactions: Actor, Partner, Relation, Activities, Context, and Evaluation. *Front. Psychol.* **2022**, *12*, 743074. [[CrossRef](#)]
40. Latour, B. On Actor-Network Theory: A Few Clarifications. *Soz. Welt* **1996**, *47*, 369–381.
41. Reeves, B.; Nass, C.I. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*; Cambridge University Press: New York, NY, USA, 1996; pp. xiv, 305. ISBN 1-57586-052-X.
42. Evans, W. 1—Mapping Virtual Worlds. In *Information Dynamics in Virtual Worlds*; Evans, W., Ed.; Chandos Information Professional Series; Chandos Publishing: Hull, UK, 2011; pp. 3–21. ISBN 978-1-84334-641-8.
43. Blumer, H. *Symbolic Interactionism: Perspective and Method*; Univ of California Press: Berkeley, CA, USA, 1986.
44. Spahn, A. Mediation in Design for Values. In *Handbook of Ethics, Values, and Technological Design*; Springer: Dordrecht, The Netherlands, 2015. [[CrossRef](#)]
45. Gagné, M.; Deci, E.L. Self-Determination Theory and Work Motivation. *J. Organ. Behav.* **2005**, *26*, 331–362. [[CrossRef](#)]
46. Shah, S.H.H.; Karlsen, A.S.T.; Solberg, M.; Hameed, I.A. A Social VR-Based Collaborative Exergame for Rehabilitation: Codesign, Development and User Study. *Virtual Real.* **2023**, *27*, 3403–3420. [[CrossRef](#)]
47. Myers, D.G. *Theories of Emotion*; Academic Press: Cambridge, MA, USA, 2004; Volume 500.
48. Lang, P. Behavioral Treatment and Bio-Behavioral Assessment: Computer Applications. *Technol. Ment. Health Care Deliv. Syst.* **1980**, *1*, 119–137.
49. Burgoon, J.K.; Buller, D.B.; Woodall, W.G. *Nonverbal Communication: The Unspoken Dialogue*; Routledge: London, UK, 1996.
50. Kinzler, K.D. Language as a Social Cue. *Annu. Rev. Psychol.* **2021**, *72*, 241–264. [[CrossRef](#)] [[PubMed](#)]
51. Le Tarnec, H.; Augereau, O.; Bevacqua, E.; De Loor, P. Impact of Augmented Engagement Model for Collaborative Avatars on a Collaborative Task in Virtual Reality. In Proceedings of the 2024 International Conference on Advanced Visual Interfaces, Genoa, Italy, 3–7 June 2024.
52. Wallkötter, S.; Tulli, S.; Castellano, G.; Paiva, A.; Chetouani, M. Explainable Embodied Agents Through Social Cues: A Review. *J. Hum.-Robot. Interact.* **2021**, *10*, 1–24. [[CrossRef](#)]

53. Sauppé, A.; Mutlu, B. How Social Cues Shape Task Coordination and Communication. In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, Baltimore, MD, USA, 15–19 February 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 97–108.
54. Feine, J.; Gnewuch, U.; Morana, S.; Maedche, A. A Taxonomy of Social Cues for Conversational Agents. *Int. J. Hum.-Comput. Stud.* **2019**, *132*, 138–161. [CrossRef]
55. Aguilar-Zambrano, J.J.; Trujillo, M.J. Factors Influencing Interaction of Creative Teams in Generation of Ideas of New Products: An Approach from Collaborative Scripts. In Proceedings of the 2017 Portland International Conference on Management of Engineering and Technology (PICMET), Portland, OR, USA, 9–13 July 2017.
56. Mallett, C.J.; Lara-Bercial, S. Chapter 14—Serial Winning Coaches: People, Vision, and Environment. In *Sport and Exercise Psychology Research*; Raab, M., Wylleman, P., Seiler, R., Elbe, A.-M., Hatzigeorgiadis, A., Eds.; Academic Press: San Diego, CA, USA, 2016; pp. 289–322. ISBN 978-0-12-803634-1.
57. Barnes, J.A. Graph Theory and Social Networks: A Technical Comment on Connectedness and Connectivity. *Sociology* **1969**, *3*, 215–232. [CrossRef]
58. Wu, Y.; You, S.; Guo, Z.; Li, X.; Zhou, G.; Gong, J. Designing a Remote Mixed-Reality Educational Game System for Promoting Children’s Social & Collaborative Skills. *arXiv* **2024**, arXiv:2301.07310.
59. IPCisco Network Topologies—•Bus•Ring•Star•Tree•Line•Mesh ★ IPCisco. Available online: <https://ipcisco.com/lesson/network-topologies/> (accessed on 19 February 2024).
60. Starzyk, K.B.; Holden, R.R.; Fabrigar, L.R.; MacDonald, T.K. The Personal Acquaintance Measure: A Tool for Appraising One’s Acquaintance with Any Person. *J. Pers. Soc. Psychol.* **2006**, *90*, 833–847. [CrossRef]
61. Zhang, G.; Zhao, S.; Liang, Z.; Li, D.; Chen, H.; Chen, X. Social Interactions With Familiar and Unfamiliar Peers in Chinese Children: Relations With Social, School, and Psychological Adjustment. *Int. Perspect. Psychol.* **2015**, *4*, 239–253. [CrossRef]
62. Mishra, N.; Schreiber, R.; Stanton, I.; Tarjan, R.E. Clustering Social Networks. In *Proceedings of the Algorithms and Models for the Web-Graph*; Bonato, A., Chung, F.R.K., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 56–67.
63. Smith, J.M. Evolution and the Theory of Games. In *Did Darwin Get It Right? Essays on Games, Sex and Evolution*; Springer: Berlin/Heidelberg, Germany, 1982; pp. 202–215.
64. Johnson, D.W.; Johnson, R.T. *Cooperation and Competition: Theory and Research*; Interaction Book Company: Edina, MN, USA, 1989; pp. viii, 253. ISBN 0-939603-10-1.
65. ADAM HAYES Game Theory. Available online: <https://www.investopedia.com/terms/g/gametheory.asp> (accessed on 8 September 2023).
66. Patrício, R.; Moreira, A.C.; Zurlo, F. Gamification in Innovation Teams. *Int. J. Innov. Stud.* **2022**, *6*, 156–168. [CrossRef]
67. Morschheuser, B.; Hamari, J.; Maedche, A. Cooperation or Competition—When Do People Contribute More? A Field Experiment on Gamification of Crowdsourcing. *Int. J. Hum.-Comput. Stud.* **2019**, *127*, 7–24. [CrossRef]
68. Giddens, A. *The Constitution of Society: Outline of the Theory of Structuration*; Univ of California Press: Berkeley, CA, USA, 1984.
69. Smith, E.R.; Semin, G.R. Socially Situated Cognition: Cognition in Its Social Context. In *Advances in Experimental Social Psychology*, Vol. 36; Elsevier Academic Press: San Diego, CA, USA, 2004; pp. 53–117. ISBN 0-12-015236-3.
70. Sawyer, K. Unresolved Tensions in Sociocultural Theory: Analogies with Contemporary Sociological Debates. *Cult. Psychol.-Cult. Psychol.* **2002**, *8*, 283–305. [CrossRef]
71. Sawyer, K. Extending Sociocultural Theory to Group Creativity. *Vocat. Learn.* **2012**, *5*, 59–75. [CrossRef]
72. Eteläpelto, A.; Vähäsantanen, K.; Hökkä, P.; Paloniemi, S. What Is Agency? Conceptualizing Professional Agency at Work. *Educ. Res. Rev.* **2013**, *10*, 45–65. [CrossRef]
73. Paliy, I.G.; Bogdanova, O.A.; Plotnikova, T.V.; Lipchanskaya, I.V. Space and Time in the Context of Social Measurement. *Eur. Res. Stud. J.* **2018**, *XXI*, 350–358. [CrossRef]
74. Klitkou, A.; Bolwig, S.; Huber, A.; Ingeborgrud, L.; Pluciński, P.; Rohracher, H.; Schartinger, D.; Thiene, M.; Żuk, P. The Interconnected Dynamics of Social Practices and Their Implications for Transformative Change: A Review. *Sustain. Prod. Consum.* **2022**, *31*, 603–614. [CrossRef]
75. Logan, J. Making a Place for Space: Spatial Thinking in Social Science. *Annu. Rev. Sociol.* **2012**, *38*, 507–524. [CrossRef] [PubMed]
76. Abbott, A. Of Time and Space: The Contemporary Relevance of the Chicago School*. *Soc. Forces* **1997**, *75*, 1149–1182. [CrossRef]
77. Ijsselsteijn, W.; Baren, J.; Lanen, F. Staying in Touch: Social Presence and Connectedness through Synchronous and Asynchronous Communication Media. *Smpte Motion Imaging J.* **2003**, *2*, e928.
78. Rauschnabel, P.A.; Felix, R.; Hinsch, C.; Shahab, H.; Alt, F. What Is XR? Towards a Framework for Augmented and Virtual Reality. *Comput. Hum. Behav.* **2022**, *133*, 107289. [CrossRef]
79. Milgram, P.; Kishino, F. A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Inf. Syst.* **1994**, *77*, 1321–1329.
80. Skarbez, R.; Smith, M.; Whitton, M.C. Revisiting Milgram and Kishino’s Reality-Virtuality Continuum. *Front. Virtual Real.* **2021**, *2*, 647997. [CrossRef]
81. Tremosa, L.; Interaction Design Foundation. Beyond AR vs. VR: What Is the Difference between AR vs. MR vs. VR vs. XR. Available online: <https://www.interaction-design.org/literature/article/beyond-ar-vs-vr-what-is-the-difference-between-ar-vs-mr-vs-vr-vs-xr> (accessed on 1 September 2024).
82. Benyon, D.; Turner, P.; Turner, S. *Designing Interactive Systems: People, Activities, Contexts, Technologies*; Pearson Education: London, UK, 2005.

83. Schleidgen, S.; Friedrich, O.; Gerlek, S.; Assadi, G.; Seifert, J. The Concept of “Interaction” in Debates on Human–Machine Interaction. *Humanit. Soc. Sci. Commun.* **2023**, *10*, 551. [CrossRef]
84. Silver, K. What Puts the Design in Interaction Design. *UX Matters* **2007**, *3*, 3–77.
85. Baykal, G.E.; Van Mechelen, M.; Goksun, T.; Yantac, A.E. Designing with and for Preschoolers: A Method to Observe Tangible Interactions with Spatial Manipulatives. In *Proceedings of the ACM International Conference Proceeding Series*; ACM Digital Library: Trondheim, Norway, 2018; pp. 45–54.
86. Bombari, D.; Schmid Mast, M.; Canadas, E.; Bachmann, M. Studying Social Interactions through Immersive Virtual Environment Technology: Virtues, Pitfalls, and Future Challenges. *Front. Psychol.* **2015**, *6*, 869. [CrossRef] [PubMed]
87. Bailenson, J.N.; Blascovich, J.; Beall, A.C.; Loomis, J.M. Interpersonal Distance in Immersive Virtual Environments. *Pers. Soc. Psychol. Bull.* **2003**, *29*, 819–833. [CrossRef]
88. Jesse Fox, K.Y.S.; Ahn, S.J.; Janssen, J.H.; Yeykelis, L.; Bailenson, J.N. Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence. *Hum.–Comput. Interact.* **2015**, *30*, 401–432. [CrossRef]
89. Nass, C.; Steuer, J.; Siminoff, E. Computer Are Social Actors. In Proceedings of the Conference on Human Factors in Computing Systems-Proceedings, Boston, MA, USA, 24–28 April 1994; p. 204.
90. Kim, D.Y.; Lee, H.K.; Chung, K. Avatar-Mediated Experience in the Metaverse: The Impact of Avatar Realism on User-Avatar Relationship. *J. Retail. Consum. Serv.* **2023**, *73*, 103382. [CrossRef]
91. Steed, A.; Schroeder, R. Collaboration in Immersive and Non-Immersive Virtual Environments. In *Immersed in Media: Telepresence Theory, Measurement & Technology*; Lombard, M., Biocca, F., Freeman, J., IJsselsteijn, W., Schaevitz, R.J., Eds.; Springer International Publishing: Cham, Germany, 2015; pp. 263–282. ISBN 978-3-319-10190-3.
92. Xiao, R.; Zhang, R.; Buruk, O.; Hamari, J.; Virkki, J. Toward Next Generation Mixed Reality Games: A Research Through Design Approach. *Prepr. Version 1 Available Res. Sq.* **2023**, *28*, 142. [CrossRef]
93. Nilsson, N.C.; Nordahl, R.; Serafin, S. Immersion Revisited: A Review of Existing Definitions of Immersion and Their Relation to Different Theories of Presence. *Hum. Technol.* **2016**, *12*, 108–134. [CrossRef]
94. Adams, E. *Fundamentals of Game Design*, 3rd ed.; New Riders: Hoboken, NJ, USA, 2006.
95. Ryan, M.-L. Interactive Narrative, Plot Types, and Interpersonal Relations. In *Proceedings of the Interactive Storytelling*; Spierling, U., Szilas, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 6–13.
96. Almeida, L.; Menezes, P.; Dias, J. Telepresence Social Robotics towards Co-Presence: A Review. *Appl. Sci.* **2022**, *12*, 5557. [CrossRef]
97. Foglia, L.; Wilson, R.A. Embodied Cognition. *WIREs Cogn. Sci.* **2013**, *4*, 319–325. [CrossRef]
98. Chao, G. Human-Computer Interaction: Process and Principles of Human-Computer Interface Design. In Proceedings of the 2009 International Conference on Computer and Automation Engineering, Bangkok, Thailand, 8–10 March 2009; pp. 230–233.
99. Nie, K.; Guo, M.; Gao, Z. Enhancing Emotional Engagement in Virtual Reality (VR) Cinematic Experiences through Multi-Sensory Interaction Design. In Proceedings of the 2023 Asia Conference on Cognitive Engineering and Intelligent Interaction (CEII), Hong Kong, China, 15–16 December 2023; pp. 47–53.
100. Broll, W.; Grimm, P.; Herold, R.; Reiners, D.; Cruz-Neira, C. VR/AR Output Devices. In *Virtual and Augmented Reality (VR/AR): Foundations and Methods of Extended Realities (XR)*; Doerner, R., Broll, W., Grimm, P., Jung, B., Eds.; Springer International Publishing: Cham, Germany, 2022; pp. 149–200. ISBN 978-3-030-79062-2.
101. Bordegoni, M.; Carulli, M.; Spadoni, E. Multisensory Interaction in eXtended Reality. In *Prototyping User eXperience in eXtended Reality*; Bordegoni, M., Carulli, M., Spadoni, E., Eds.; Springer Nature Switzerland: Cham, Germany, 2023; pp. 49–63. ISBN 978-3-031-39683-0.
102. Verbeek, P.-P. COVER STORYBeyond Interaction: A Short Introduction to Mediation Theory. *Interactions* **2015**, *22*, 26–31. [CrossRef]
103. Ihde, D. *Technology and the Lifeworld: From Garden to Earth*; Indiana University Press: Bloomington, IN, USA, 1990.
104. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann. Intern. Med.* **2009**, *151*, 264–269. [CrossRef]
105. Scopus—Document Search | Signed In. Available online: <https://www.scopus.com/search/form.uri?display=basic#basic> (accessed on 18 October 2023).
106. Brown, K.E.; Heise, N.; Eitel, C.M.; Nelson, J.; Garbe, B.A.; Meyer, C.A.; Ivie, K.R.; Clapp, T.R. A Large-Scale, Multiplayer Virtual Reality Deployment: A Novel Approach to Distance Education in Human Anatomy. *Med. Sci. Educ.* **2023**, *33*, 409–421. [CrossRef]
107. Bird, J.M.; Smart, P.A.; Harris, D.J.; Phillips, L.A.; Giannachi, G.; Vine, S.J. A Magic Leap in Tourism: Intended and Realized Experience of Head-Mounted Augmented Reality in a Museum Context. *J. Travel Res.* **2023**, *62*, 1427–1447. [CrossRef]
108. Healey, J.; Wang, D.; Wigington, C.; Sun, T.; Peng, H. A Mixed-Reality System to Promote Child Engagement in Remote Intergenerational Storytelling. In Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Bari, Italy, 4–8 October 2021; Volume 2021, pp. 274–279. [CrossRef]
109. Vella, K.; Johnson, D.; Cheng, V.W.S.; Davenport, T.; Mitchell, J.; Klarkowski, M.; Phillips, C. A Sense of Belonging: Pokémon GO and Social Connectedness. *Games Cult.* **2019**, *14*, 583–603. [CrossRef]
110. Ballestin, G.; Bassano, C.; Solari, F.; Chessa, M. A Virtual Reality Game Design for Collaborative Team-Building: A Proof of Concept. In Proceedings of the UMAP ‘20: 28th ACM Conference on User Modeling, Adaptation and Personalization, Genoa, Italy, 14–17 July 2020; pp. 159–162. [CrossRef]

111. Cassidy, B.; Sim, G.; Robinson, D.W.; Gandy, D. A Virtual Reality Platform for Analyzing Remote Archaeological Sites. *Interact. Comput.* **2019**, *31*, 167–176. [[CrossRef](#)]
112. Maloney, D.; Freeman, G.; Robb, A. A Virtual Space for All: Exploring Children’s Experience in Social Virtual Reality. In Proceedings of the CHI PLAY’20: The Annual Symposium on Computer-Human Interaction in Play, Virtual Event, Canada, 2–4 November 2020; pp. 472–483. [[CrossRef](#)]
113. Koukopoulos, Z.; Koukopoulos, D. Active Visitor: Augmenting Libraries into Social Spaces. In Proceedings of the 2018 3rd Digital Heritage International Congress (DigitalHERITAGE) Held Jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018), San Francisco, CA, USA, 26–30 October 2018. [[CrossRef](#)]
114. Jung, S.; Wu, Y.; McKee, R.; Lindeman, R.W. All Shook Up: The Impact of Floor Vibration in Symmetric and Asymmetric Immersive Multi-User VR Gaming Experiences. In Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Christchurch, New Zealand, 12–16 March 2022; pp. 737–745. [[CrossRef](#)]
115. Gillespie, D.; Qin, Z.; Aish, F. An Extended Reality Collaborative Design System: In-Situ Design Reviews in Uncontrolled Environments. In Proceedings of the 2021 ACADIA, Association for Computer-Aided Design in Architecture, Conference, Calgary, AB, Canada, 3–6 November 2021.
116. Echavarria, K.R.; Dibble, L.; Bracco, A.; Silverton, E.; Dixon, S. Augmented Reality (AR) Maps for Experiencing Creative Narratives of Cultural Heritage. In *EUROGRAPHICS Workshop on Graphics and Cultural Heritage*; The Eurographics Association: Eindhoven, The Netherlands, 2019; pp. 7–16. [[CrossRef](#)]
117. Jiang, S.; Moyle, B.; Yung, R.; Tao, L.; Scott, N. Augmented Reality and the Enhancement of Memorable Tourism Experiences at Heritage Sites. *Curr. Issues Tour.* **2023**, *26*, 242–257. [[CrossRef](#)]
118. Ahmadi Oloonabadi, S.; Baran, P. Augmented Reality Participatory Platform: A Novel Digital Participatory Planning Tool to Engage under-Resourced Communities in Improving Neighborhood Walkability. *Cities* **2023**, *141*, 104441. [[CrossRef](#)]
119. Prandi, C.; Nisi, V.; Ceccarini, C.; Nunes, N. Augmenting Emerging Hospitality Services: A Playful Immersive Experience to Foster Interactions among Locals and Visitors. *Int. J. Hum.-Comput. Interact.* **2023**, *39*, 363–377. [[CrossRef](#)]
120. Wang, Y.; Gardner, H.; Martin, C.; Adcock, M. Augmenting Sculpture with Immersive Sonification. In Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Christchurch, New Zealand, 12–16 March 2022; pp. 626–627. [[CrossRef](#)]
121. Kohler, T.; Fueller, J.; Stieger, D.; Matzler, K. Avatar-Based Innovation: Consequences of the Virtual Co-Creation Experience. *Comput. Hum. Behav.* **2011**, *27*, 160–168. [[CrossRef](#)]
122. Xu, Y.; Gandy, M.; Deen, S.; Schrank, B.; Spreen, K.; Gorbsky, M.; White, T.; Barba, E.; Radu, I.; Bolter, J.; et al. BragFish: Exploring Physical and Social Interaction in Co-Located Handheld Augmented Reality Games. In Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology, Yokohama, Japan, 3–5 December 2008; Association for Computing Machinery: New York, NY, USA, 2008; pp. 276–283.
123. Samayoa, A.G.; Talavera, J.; Sium, S.G.; Xie, B.; Huang, H.; Yu, L.-F. Building a Motion-Aware, Networked Do-It-Yourself Holographic Display. In Proceedings of the 2021 IEEE International Conference on Intelligent Reality (ICIR), Virtual Conference, 12–13 May 2021; pp. 39–48. [[CrossRef](#)]
124. Yavo-Ayalon, S.; Joshi, S.; Zhang, Y.; Han, R.; Mahyar, N.; Ju, W. Building Community Resiliency through Immersive Communal Extended Reality (CXR). *Multimodal Technol. Interact.* **2023**, *7*, 43. [[CrossRef](#)]
125. Faridan, M.; Kumari, B.; Suzuki, R. ChameleonControl: Teleoperating Real Human Surrogates through Mixed Reality Gestural Guidance for Remote Hands-on Classrooms. In Proceedings of the CHI’23: CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023. [[CrossRef](#)]
126. Chung, C.-Y.; Awad, N.; Hsiao, H. Collaborative Programming Problem-Solving in Augmented Reality: Multimodal Analysis of Effectiveness and Group Collaboration. *Australas. J. Educ. Technol.* **2021**, *37*, 17–31. [[CrossRef](#)]
127. Feng, Y.; Perugia, G.; Yu, S.; Barakova, E.I.; Hu, J.; Rauterberg, G.W.M. Context-Enhanced Human-Robot Interaction: Exploring the Role of System Interactivity and Multimodal Stimuli on the Engagement of People with Dementia. *Int. J. Soc. Robot.* **2022**, *14*, 807–826. [[CrossRef](#)]
128. Frydenberg, M.; Andone, D. Converging Digital Literacy through Virtual Reality. In Proceedings of the 2021 IEEE Frontiers in Education Conference (FIE), Lincoln, NE, USA, 13–16 October 2021. [[CrossRef](#)]
129. Neate, T.; Roper, A.; Wilson, S.; Marshall, J.; Cruice, M. CreaTable Content and Tangible Interaction in Aphasia. In Proceedings of the Conference on Human Factors in Computing Systems-Proceedings, Honolulu, HI, USA, 23 April 2020; pp. 1–14.
130. Spiel, K.K.; Werner, K.; Hödl, O.; Ehrenstrasser, L.; Fitzpatrick, G. Creating Community Fountains by (Re-)Designing the Digital Layer of Way-Finding Pillars. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, Vienna, Austria, 4–7 September 2017. [[CrossRef](#)]
131. Li, Y.; Yu, L.; Liang, H.-N. CubeMuseum: An Augmented Reality Prototype of Embodied Virtual Museum. In Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Bari, Italy, 4–8 October 2021; pp. 13–17. [[CrossRef](#)]
132. Kim, M.; Lee, S.-H. Deictic Gesture Retargeting for Telepresence Avatars in Dissimilar Object and User Arrangements. In Proceedings of the Web3D’20: The 25th International Conference on 3D Web Technology, Virtual Event, Republic of Korea, 9–13 November 2020. [[CrossRef](#)]

133. Costa, M.C.; Manso, A.; Patrício, J. Design of a Mobile Augmented Reality Platform with Game-Based Learning Purposes. *Information* **2020**, *11*, 127. [[CrossRef](#)]
134. Xu, T.B.; Mostafavi, A.; Kim, B.C.; Lee, A.A.; Boot, W.; Czaja, S.; Kalantari, S. Designing Virtual Environments for Social Engagement in Older Adults: A Qualitative Multi-Site Study. In Proceedings of the CHI'23: CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023. [[CrossRef](#)]
135. Oksman, V.; Kulju, M. Developing Online Illustrative and Participatory Tools for Urban Planning: Towards Open Innovation and Co-Production through Citizen Engagement. *Int. J. Serv. Technol. Manag.* **2017**, *23*, 445–464. [[CrossRef](#)]
136. Siriaraya, P.; Ang, C.S. Developing Virtual Environments for Older Users: Case Studies of Virtual Environments Iteratively Developed for Older Users and People with Dementia. In Proceedings of the 2017 2nd International Conference on Information Technology (INCIT), Nakhonpathom, Thailand, 2–3 November 2017; pp. 1–6. [[CrossRef](#)]
137. Schmitt, L.; Buisine, S.; Chaboissier, J.; Aoussat, A.; Vernier, F. Dynamic Tabletop Interfaces for Increasing Creativity. *Comput. Hum. Behav.* **2012**, *28*, 1892–1901. [[CrossRef](#)]
138. Pokric, B.; Krco, S.; Pokric, M.; Knezevic, P.; Jovanovic, D. Engaging Citizen Communities in Smart Cities Using IoT, Serious Gaming and Fast Markerless Augmented Reality. In Proceedings of the 2015 International Conference on Recent Advances in Internet of Things (RIoT), Singapore, 7–9 April 2015. [[CrossRef](#)]
139. Erzetic, C.; Dobbs, T.; Fabbri, A.; Gardner, N.; Haeusler, M.H.; Zavoleas, Y. Enhancing User-Engagement in the Design Process through Augmented Reality Applications. *Proc. Int. Conf. Educ. Res. Comput. Aided Archit. Des. Eur.* **2019**, *2*, 423–432. [[CrossRef](#)]
140. Li, J.; Van Der Spek, E.D.; Yu, X.; Hu, J.; Feijis, L. Exploring an Augmented Reality Social Learning Game for Elementary School Students. In Proceedings of the interaction design and children conference, London, UK, 21–24 June 2020; pp. 508–518. [[CrossRef](#)]
141. McCaffery, J.; Miller, A.; Kennedy, S.; Dawson, T.; Allison, C.; Vermehren, A.; Lefley, C.; Strickland, K. Exploring Heritage through Time and Space Supporting Community Reflection on the Highland Clearances. In Proceedings of the 2013 Digital Heritage International Congress (DigitalHeritage), Marseille, France, 28 October–1 November 2013; Volume 1, pp. 371–378. [[CrossRef](#)]
142. Kersting, M.; Steier, R.; Venville, G. Exploring Participant Engagement during an Astrophysics Virtual Reality Experience at a Science Festival. *Int. J. Sci. Educ. Part B Commun. Public Engagem.* **2021**, *11*, 17–34. [[CrossRef](#)]
143. Gruber, A.; Canto, S.; Jauregi-Ondarra, K. Exploring the Use of Social Virtual Reality for Virtual Exchange. *ReCALL* **2023**, *35*, 258–273. [[CrossRef](#)]
144. Gugenheimer, J.; Stemasov, E.; Sareen, H.; Rukzio, E. FaceDisplay: Towards Asymmetric Multi-User Interaction for Nomadic Virtual Reality. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–13.
145. DiBenigno, M.; Kosa, M.; Johnson-Glenberg, M.C. Flow Immersive: A Multiuser, Multidimensional, Multiplatform Interactive Covid-19 Data Visualization Tool. *Front. Psychol.* **2021**, *12*, 661613. [[CrossRef](#)]
146. Vets, T.; Nijs, L.; Lesaffre, M.; Moens, B.; Bressan, F.; Colpaert, P.; Lambert, P.; Van de Walle, R.; Leman, M. Gamified Music Improvisation with BilliArT: A Multimodal Installation with Balls. *J. Multimodal User Interfaces* **2017**, *11*, 25–38. [[CrossRef](#)]
147. Planey, J.; Rajarathinam, R.J.; Mercier, E.; Lindgren, R. Gesture-Mediated Collaboration with Augmented Reality Headsets in a Problem-Based Astronomy Task. *Int. J. Comput.-Support. Collab. Learn.* **2023**, *18*, 259–289. [[CrossRef](#)]
148. Williams, S.; Enatsky, R.; Gillcash, H.; Murphy, J.J.; Gracanin, D. Immersive Technology in the Public School Classroom: When a Class Meets. In Proceedings of the 2021 7th International Conference of the Immersive Learning Research Network (iLRN), Eureka, CA, USA, 17 May–10 June 2021. [[CrossRef](#)]
149. Pereira, F.; Bermudez I Badia, S.; Jorge, C.; Da Silva Cameirao, M. Impact of Game Mode on Engagement and Social Involvement in Multi-User Serious Games with Stroke Patients. In Proceedings of the 2019 international conference on virtual rehabilitation (ICVR), Tel Aviv, Israel, 21–24 July 2019. [[CrossRef](#)]
150. Doukianou, S.; Daylamani-Zad, D.; O’Loingsigh, K. Implementing an Augmented Reality and Animated Infographics Application for Presentations: Effect on Audience Engagement and Efficacy of Communication. *Multimed. Tools Appl.* **2021**, *80*, 30969–30991. [[CrossRef](#)]
151. Jackson, D.; Kaveh, H.; Victoria, J.; Walker, A.; Bursztyn, N. Integrating an Augmented Reality Sandbox Challenge Activity into a Large-Enrollment Introductory Geoscience Lab for Nonmajors Produces No Learning Gains. *J. Geosci. Educ.* **2019**, *67*, 237–248. [[CrossRef](#)]
152. Postert, P.; Wolf, A.E.M.; Schiewe, J. Integrating Visualization and Interaction Tools for Enhancing Collaboration in Different Public Participation Settings. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 156. [[CrossRef](#)]
153. Pedram, S.; Palmisano, S.; Skarbez, R.; Perez, P.; Farrelly, M. Investigating the Process of Mine Rescuers’ Safety Training with Immersive Virtual Reality: A Structural Equation Modelling Approach. *Comput. Educ.* **2020**, *153*, 103891. [[CrossRef](#)]
154. Papadimitriou, S.; Kamitsios, M.; Chrysafiadi, K.; Virvou, M. Learn-and-Play Personalised Reasoning from Point-and-Click to Virtual Reality Mobile Educational Games. *Intell. Decis. Technol.* **2021**, *15*, 321–332. [[CrossRef](#)]
155. Vayanou, M.; Christodoulou, K.; Katifori, A.; Ioannidis, Y. MagicARTs: On the Design of Social VR Experiences. *CEUR Workshop Proc.* **2022**, *3243*.
156. Sayis, B.; Ramirez, R.; Pares, N. Mixed Reality or LEGO Game Play? Fostering Social Interaction in Children with Autism. *Virtual Real.* **2022**, *26*, 771–787. [[CrossRef](#)]

157. Wu, Y.; You, S.; Guo, Z.; Li, X.; Zhou, G.; Gong, J.M.R. Brick: Designing A Remote Mixed-Reality Educational Game System for Promoting Children's Social & Collaborative Skills. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 1–18. [[CrossRef](#)]
158. Tsoupikova, D.; Triandafilou, K.; Thielbar, K.; Rupp, G.; Preuss, F.; Kamper, D. Multi-User Virtual Reality Therapy for Post-Stroke Hand Rehabilitation at Home. *J. Syst. Cybernet. Informat.* **2016**, *1*, 227–231.
159. Drljević, N.; Botički, I.; Wong, L.H. Observing Student Engagement during Augmented Reality Learning in Early Primary School. *J. Comput. Educ.* **2022**, *11*, 181–213. [[CrossRef](#)]
160. Irie, K.; Sada, M.A.; Yamada, Y.; Gushima, K.; Nakajima, T. Pervasive HoloMoL: A Mobile Pervasive Game with Mixed Reality Enhanced Method of Loci. In Proceedings of the 15th International Conference on Advances in Mobile Computing & Multimedia, Salzburg, Austria, 4–6 December 2017; pp. 141–145. [[CrossRef](#)]
161. Basballe, D.A.; Halskov, K. Projections on Museum Exhibits—Engaging Visitors in the Museum Setting. In Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction, Brisbane, Australia, 22–26 November 2010; pp. 80–87. [[CrossRef](#)]
162. Wang, H.-Y.; Sun, J.C.-Y. Real-Time Virtual Reality Co-Creation: Collective Intelligence and Consciousness for Student Engagement and Focused Attention within Online Communities. *Interact. Learn. Environ.* **2023**, *31*, 3422–3435. [[CrossRef](#)]
163. Lee, M.-C. Rediscovering Neighborhood History Through Augmented Reality. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Taichung, Taiwan, 15–17 November 2021; pp. 60–64. [[CrossRef](#)]
164. Fan, J.; Mion, L.C.; Beuscher, L.; Ullal, A.; Newhouse, P.A.; Sarkar, N. SAR-Connect: A Socially Assistive Robotic System to Support Activity and Social Engagement of Older Adults. *IEEE Trans. Robot.* **2022**, *38*, 1250–1269. [[CrossRef](#)]
165. Puggioni, M.; Frontoni, E.; Paolanti, M.; Pierdicca, R. ScoolAR: An Educational Platform to Improve Students' Learning through Virtual Reality. *IEEE Access* **2021**, *9*, 21059–21070. [[CrossRef](#)]
166. Yang, K.-T.; Wang, C.-H.; Chan, L. ShareSpace: Facilitating Shared Use of the Physical Space by both VR Head-Mounted Display and External Users. In Proceedings of the UIST’18: The 31st Annual ACM Symposium on User Interface Software and Technology, Berlin, Germany, 14 October 2018; pp. 499–509. [[CrossRef](#)]
167. Gugenheimer, J.; Stemasov, E.; Frommel, J.; Rukzio, E. ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 4021–4033.
168. Jing, A.; Xiang, C.; Kim, S.; Billinghurst, M.; Quigley, A. SnapChart: An Augmented Reality Analytics Toolkit to Enhance Interactivity in a Collaborative Environment. In Proceedings of the VRCAI’19: The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry, Brisbane, Australia, 14–16 November 2019. [[CrossRef](#)]
169. Dodoo, N.A.; Youn, S. Snapping and Chatting Away: Consumer Motivations for and Outcomes of Interacting with Snapchat AR Ad Lens. *Telemat. Inform.* **2021**, *57*, 101514. [[CrossRef](#)]
170. Snibbe, S.S.; Raffle, H.S. Social Immersive Media: Pursuing Best Practices for Multi-User Interactive Camera/Projector Exhibits. In Proceedings of the CHI’09: CHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 4–9 April 2009; pp. 1447–1456. [[CrossRef](#)]
171. Wagner, I.; Basile, M.; Ehrenstrasser, L.; Maquil, V.; Terrin, J.-J.; Wagner, M. Supporting Community Engagement in the City: Urban Planning in the MR-Tent. In Proceedings of the C&T’09: Communities and Technologies, University Park, PA, USA, 25–27 June 2009; pp. 185–194. [[CrossRef](#)]
172. Norman, M.; Lee, G.A.; Smith, R.T.; Billinghurst, M. The Impact of Remote User’s Role in a Mixed Reality Mixed Presence System. In Proceedings of the VRCAI’19: The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry, Brisbane, Australia, 14–16 November 2019. [[CrossRef](#)]
173. Bekele, M.K.; Champion, E.; McMeekin, D.A.; Rahaman, H. The Influence of Collaborative and Multi-Modal Mixed Reality: Cultural Learning in Virtual Heritage. *Multimodal Technol. Interact.* **2021**, *5*, 79. [[CrossRef](#)]
174. Cheok, A.D.; Yang, X.; Ying, Z.Z.; Billinghurst, M.; Kato, H. Touch-Space: Mixed Reality Game Space Based on Ubiquitous, Tangible, and Social Computing. *Pers. Ubiquitous Comput.* **2002**, *6*, 430–442. [[CrossRef](#)]
175. Moreira, C.; Simoes, F.P.M.; Lee, M.J.W.; Zorzal, E.R.; Lindeman, R.W.; Pereira, J.M.; Johnsen, K.; Jorge, J. Toward VR in VR: Assessing Engagement and Social Interaction in a Virtual Conference. *IEEE Access* **2023**, *11*, 1906–1922. [[CrossRef](#)]
176. Wagner, D.; Pintaric, T.; Ledermann, F.; Schmalstieg, D. Towards Massively Multi-User Augmented Reality on Handheld Devices. *Lect. Notes Comput. Sci.* **2005**, *3468*, 208–219. [[CrossRef](#)]
177. Kalantari, S.; Xu, T.B.; Mostafavi, A.; Kim, B.; Dilanchian, A.; Lee, A.; Boot, W.R.; Czaja, S.J. Using Immersive Virtual Reality to Enhance Social Interaction Among Older Adults: A Cross-Site Investigation. *Innov. Aging* **2023**, *7*, igad031. [[CrossRef](#)]
178. Meenar, M.; Kitson, J. Using Multi-Sensory and Multi-Dimensional Immersive Virtual Reality in Participatory Planning. *Urban Sci.* **2020**, *4*, 34. [[CrossRef](#)]
179. Taylor, M.; Kaur, M.; Sharma, U.; Taylor, D.; Reed, J.E.; Darzi, A. Using Virtual Worlds for Patient and Public Engagement. *Int. J. Technol. Knowl. Soc.* **2013**, *9*, 31–48. [[CrossRef](#)] [[PubMed](#)]
180. Pimentel, D.; Kalyanaraman, S. Virtual Climate Scientist: A VR Learning Experience about Paleoclimatology for Underrepresented Students. *Interact. Learn. Environ.* **2023**, *31*, 4426–4439. [[CrossRef](#)]

181. Brůža, V.; Byška, J.; Mičan, J.; Kozlíková, B. VRdeo: Creating Engaging Educational Material for Asynchronous Student-Teacher Exchange Using Virtual Reality. *Comput. Graph. Pergamon* **2021**, *98*, 280–292. [CrossRef]
182. Liaw, S.Y.; Choo, T.; Wu, L.T.; Lim, W.S.; Choo, H.; Lim, S.M.; Ringsted, C.; Wong, L.F.; Ooi, S.L.; Lau, T.C. “Wow, Woo, Win”: Healthcare Students’ and Facilitators’ Experiences of Interprofessional Simulation in Three-Dimensional Virtual World: A Qualitative Evaluation Study. *Nurse Educ. Today* **2021**, *105*, 105018. [CrossRef]
183. Wu, Y.; Wang, Y.; Lou, X. A Large Display-Based Approach Supporting Natural User Interaction in Virtual Reality Environment. *Int. J. Ind. Ergon.* **2024**, *101*, 103591. [CrossRef]
184. Flynn, A.; Koh, W.Q.; Reilly, G.; Brennan, A.; Redfern, S.; Barry, M.; Casey, D. A Multi-User Virtual Reality Social Connecting Space for People Living with Dementia and Their Support Persons: A Participatory Action Research Study. *Int. J. Hum.-Comput. Interact.* **2024**, *40*, 1–19. [CrossRef]
185. Ahmadpour, N.; Pillai, A.G.; Yao, S.; Weatherall, A. Building Enriching Realities with Children: Creating Makerspaces That Intertwine Virtual and Physical Worlds in Pediatric Hospitals. *Int. J. Hum. Comput. Stud.* **2024**, *183*, 103193. [CrossRef]
186. Shadiev, R.; Chen, X.; Reynolds, B.L.; Song, Y.; Altinay, F. Facilitating Cognitive Development and Addressing Stereotypes with a Cross-Cultural Learning Activity Supported by Interactive 360-Degree Video Technology. *Br. J. Educ. Technol.* **2024**. [CrossRef]
187. Chen, Y.-T.; Li, M.; Cukurova, M.; Jong, M.S.-Y. Incorporation of Peer-Feedback into the Pedagogical Use of Spherical Video-Based Virtual Reality in Writing Education. *Br. J. Educ. Technol.* **2024**, *55*, 519–540. [CrossRef]
188. DeVeaux, C.; Markowitz, D.M.; Han, E.; Miller, M.R.; Hancock, J.T.; Bailenson, J.N. Presence and Pronouns: An Exploratory Investigation into the Language of Social VR. *J. Lang. Soc. Psychol.* **2024**, *43*, 405–427. [CrossRef]
189. Boubakri, F.-E.; Kadri, M.; Kaghat, F.Z.; Azough, A. Virtual Reality Classrooms vs. Video Conferencing Platform, Initial Design and Evaluation Study for Collaborative Distance Learning. *Multimed. Tools Appl.* **2024**, *83*, 1–32. [CrossRef]
190. Do, N.H.; Le, K.-D.; Fjeld, M.; Ly, D.-N.; Tran, M.-T. XRPublicSpectator: Towards Public Mixed Reality Viewing in Collocated Asymmetric Groups. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024.
191. Remy, C.; MacDonald Vermeulen, L.; Frich, J.; Biskjaer, M.M.; Dalsgaard, P. Evaluating Creativity Support Tools in HCI Research. In Proceedings of the 2020 ACM Designing Interactive Systems Conference, Eindhoven, The Netherlands, 6–10 July 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 457–476.
192. Jiang, M.; Wang, Y.; Nanjappan, V.; Bai, Z.; Liang, H.-N. E-Textiles for Emotion Interaction: A Scoping Review of Trends and Opportunities. *Pers. Ubiquitous Comput.* **2024**, *11*, 1–29. [CrossRef]
193. Brandes, U. A Faster Algorithm for Betweenness Centrality*. *J. Math. Sociol.* **2001**, *25*, 163–177. [CrossRef]
194. Kyrlitsias, C.; Michael-Grigoriou, D. Social Interaction With Agents and Avatars in Immersive Virtual Environments: A Survey. *Front. Virtual Real.* **2022**, *2*, 786665. [CrossRef]
195. Shapiro, L.; Spaulding, S. Embodied Cognition. In *The Stanford Encyclopedia of Philosophy*; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2021.
196. Walther, J. Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal. *Commun. Res.* **1996**, *23*, 3–43. [CrossRef]
197. Wang, Y.; Dai, Y.; Chen, S.; Wang, L.; Hoorn, J.F. Multiplayer Online Battle Arena (MOBA) Games: Improving Negative Atmosphere with Social Robots and AI Teammates. *Systems* **2023**, *11*, 425. [CrossRef]
198. Qin, H.X.; Hui, P. Empowering the Metaverse with Generative AI: Survey and Future Directions. In Proceedings of the 2023 IEEE 43rd International Conference on Distributed Computing Systems Workshops (ICDCSW), Hong Kong, China, 18–21 July 2023; pp. 85–90.
199. Sørensen, B.; Thellefsen, M.; Thellefsen, T. Assistive Technologies and Habit Development: A Semiotic Model of Technological Mediation. *Lang. Semiot. Stud.* **2024**, *10*, 43–64. [CrossRef]
200. Moon VR Home Enter Virtual Worlds of Your Wish 2024. Available online: <https://moonvrhome.com/> (accessed on 1 September 2024).
201. Segovia, K.Y.; Bailenson, J.N. Virtually True: Children’s Acquisition of False Memories in Virtual Reality. *Media Psychol.* **2009**, *12*, 371–393. [CrossRef]
202. Lee, S.; El Ali, A.; Wijntjes, M.; Cesar, P. Understanding and Designing Avatar Biosignal Visualizations for Social Virtual Reality Entertainment. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 30 April–6 May 2022; Association for Computing Machinery: New York, NY, USA, 2022.
203. Curran, M.T.; Gordon, J.R.; Lin, L.; Sridhar, P.K.; Chuang, J. Understanding Digitally-Mediated Empathy: An Exploration of Visual, Narrative, and Biosensory Informational Cues. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Scotland, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–13.
204. Aoudia, F.A.; Hoydis, J. Model-Free Training of End-to-End Communication Systems. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2503–2516. [CrossRef]