



A survey of challenges and methods for Quality of Experience assessment of interactive VR applications

Sara Vlahovic¹ · Mirko Suznjevic¹ · Lea Skorin-Kapov¹

Received: 1 June 2021 / Accepted: 24 March 2022 / Published online: 29 April 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

User acceptance of virtual reality (VR) applications is dependent on multiple aspects, such as usability, enjoyment, and cybersickness. To fully realize the disruptive potential of VR technology in light of recent technological advancements (e.g., advanced headsets, immersive graphics), gaining a deeper understanding of underlying factors and dimensions impacting and contributing to the overall end-user experience is of great benefit to hardware manufacturers, software and content developers, and service providers. To provide insight into user behaviour and preferences, researchers conduct user studies exploring the influence of various user-, system-, and context-related factors on the overall Quality of Experience (QoE) and its dimensions. When planning and executing such studies, researchers are faced with numerous methodological challenges related to study design aspects, such as specification of dependant and independent variables, subjective and objective assessment methods, preparation of test materials, test environment, and participant recruitment. Approaching these challenges from a multidisciplinary perspective, this paper reviews different aspects of performing perception-based QoE assessment for interactive VR applications and presents options and recommendations for research methodology design. We provide an overview of different influence factors and dimensions that may affect the overall QoE, with a focus on presence, immersion, and discomfort. Furthermore, we address ethical and practical issues regarding participant choice and test material, present different assessment methods and measures commonly used in VR research, and discuss approaches to choosing study duration and location. Lastly, we provide a concise analysis of key challenges that need to be addressed in future studies centered around VR QoE.

Keywords Quality of Experience · Virtual Reality · User Research · Immersive Technology

1 Introduction

In the year 2011 the first prototype of Oculus Rift was designed, marking the beginning of a new era of virtual reality (VR). Over the next several years, Oculus Rift was joined by similar VR systems, the most notable being HTC Vive, PlayStation VR, and Valve Index. Unlike most devices that came before, mid-2010s commercial VR systems showcased high-quality features whilst remaining within a relatively

affordable price range. The public was intrigued by the promises of unprecedented levels of immersion and novel ways of interacting with the virtual world. In fact, the results of recent studies confirm what is considered to be VR's selling point—not only is VR more appealing [1] compared to non-immersive platforms, but playing a game in VR appears to improve perceived presence [2], as well as the overall satisfaction [3], enjoyment [3], and happiness [2]. And yet, as of March 2021—a decade after the initial Oculus Rift design—the percentage of VR headset owners among Steam users sits at a fairly low 2.60%, with only a slight increase of 0.09% compared to the previous month [4].

To fully realize the disruptive potential of immersive technologies such as VR, as well as augmented reality (AR), hardware manufacturers and content developers need to address the issue from the perspective of the end user. Based on the definition given in the Qualinet White Paper on Definitions of Quality of Experience [5], QoE for immersive

✉ Sara Vlahovic
sara.vlahovic@fer.hr

Mirko Suznjevic
mirko.suznjevic@fer.hr

Lea Skorin-Kapov
lea.skorin-kapov@fer.hr

¹ Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia

media, such as VR, is defined as: “*the degree of delight or annoyance of the user of an application or service which involves an immersive media experience. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user’s personality and current state.*” [6]. This paper brings together key findings from literature and standards related to QoE assessment for VR, and provides readers with a systematic overview and guidelines on key aspects to be considered when planning and executing QoE assessment studies. While many individual aspects addressed in this paper have already been discussed at length in various research publications (e.g., [7–9]), our paper is conceived as a comprehensive collection of valuable resources, recommendations and explanations of relevant concepts, tools and methods, to be used as a reference for scientists and industry experts looking to incorporate user testing into their research and development process. As such, it provides information that may be useful to researchers from various disciplines and presents methods suitable for testing interactive VR services regardless of their intended use.

Parés and Parés [10] discuss the distinction between the terms virtual environment (VE) and virtual reality (VR). According to their definition, VE refers to a static environment comprised of different content, geometry and static rules of the environment, while VR refers to a VE in action (i.e., experienced in real time). However, this paper will refer to the term virtual reality as defined by Aukstakalnis [11], who refers to VR as different display technologies (e.g., head-mounted display—HMD, computer-assisted virtual environments—CAVE) capable of generating sensations of immersion and presence inside a three-dimensional model or simulation, therefore creating a visual replacement of the real world. As such, VR belongs to what we refer to as Immersive Media Technologies (IMT). State of the art information on IMT and Immersive Media Experience (IMEx) has recently been presented in the QUALINET White Paper on Definitions of Immersive Media Experience (IMEx) [6]. A systematic literature review on the topic of immersive systems is provided by Liberatore and Wagner [12].

While VR systems can be used for viewing 360-degree videos (e.g., [13]), we highlight that this paper focuses on QoE for interactive applications. Considering that interactivity is defined as “*the extent to which users can participate in modifying the form and content of a mediated environment in real time*” [14], we consider interactive VR applications as being those which enable users to navigate and/or manipulate the virtual environment, instead of passively observing. Interactive VR applications share many similarities with conventional interactive applications, such as computer games, which means that papers and recommendations related to gaming QoE may serve as useful guidelines for designing VR applications and conducting VR studies. While further

focusing on immersion and presence, what distinguishes VR from non-immersive platforms are issues of physical side effects (i.e., cybersickness) and general discomfort, which plague as much as 80% of users [15] and therefore present a serious obstacle to achieving not only high levels of user satisfaction, but, more importantly, providing a healthy and safe experience. Unfortunately, optimizing these factors proves to be especially problematic as presence and cybersickness appear to be negatively correlated [16]—by making application design choices aimed at improving perceived presence, developers are more likely to provoke cybersickness symptoms in users. This paradox highlights the importance of close examination of various factors that contribute to increasing perceived presence and decreasing cybersickness, as well as the importance of a thorough analysis of ways in which these factors relate to each other and contribute to the overall QoE score.

This paper focuses on perception-based methods of QoE assessment, i.e., methods based on testing human evaluators (participants). As explained in [17], participants in user studies may be presented with a test stimulus or multiple test stimuli, asked to interact with a system, and/or use the system in interaction with another person. Based on these experiences, users provide quantitative or qualitative subjective evaluations, which subsequently undergo statistical analysis. In addition to subjective measurements, researchers often employ objective methods of evaluation, such as different physiological, behavioral, and task performance measurements. Considering that QoE as a field of research focuses on the isolation of specific factors, perception-based QoE assessment may serve as a foundation for analysis of individual QoE elements, in addition to being a step towards ascertaining the value of QoE as a holistic concept. The overall QoE of a VR application, service, or system, can therefore be considered a combination of its elements, although their individual relationships can only be determined based on experimental data.

In this paper, we provide an analysis of different factors impacting QoE for interactive, synthetic, locally-rendered VR applications, with a special focus on aforementioned key elements of VR applications (i.e., presence, immersion, and cybersickness) and offer a comprehensive overview of published work. While the various addressed aspects of QoE assessment are applicable for both single and multi-user scenarios (referring to VR meetings and collaborative applications), we note that VR telemeetings imply an additional set of influence factors and QoE dimensions. While we do not address these aspects in detail, the interested reader is referred to the baseline draft of the ITU-T Recommendation for QoE assessment of extended reality (XR) meetings [18], developed in the scope of Study Group 12.

As we discuss the multidisciplinary field of QoE in the context of VR as a multimodal platform developed for a vari-

ety of use-cases, ranging from medicine and engineering to education and entertainment, our guiding principle in writing this paper is to adopt a thoroughly integrative approach. Considering the overarching similarities between the two (as discussed in [19]), we include notable studies and research guidelines stemming from the field of User Experience (UX) in addition to sources directly pertaining to the field of QoE. To further explain the reasoning behind discussed or proposed methodology choices (especially in relation to the issues of ethics and safety), we refer to research encompassing several disciplines, such as psychology, medicine, telecommunications, and computer science.

The structure of the paper (Sects. 2–8) loosely follows the set of questions referred to as *the seven circumstances* [20], as illustrated in Fig. 1, which presents the research questions we aim to address in each section. Using this format, we aim to systematically discuss different aspects of perception-based QoE assessment, such as different influence factors and QoE features, the knowledge of which can aid in defining the study objective, to ways in which researchers can provide a safe, ethical research environment, eliminate bias in participant choice and study design, and define appropriate study methodologies with respect to external and internal validity. Key challenges and ideas for future work are presented in Sect. 9, while our concluding remarks are presented in Sect. 10.

2 The importance of QoE assessment for immersive VR applications

According to a survey conducted by Perkins Coie for the year 2020 [21], limited quality and/or quantity of available VR content is considered to be the biggest obstacle to mass adoption of VR, followed by inadequate user experience, and consumer and business reluctance to use AR/VR technology. These obstacles appear to be interconnected—e.g., if the goal is to encourage customers to increase the “demand” for content, developers and manufacturers would first have to improve the quality of the “supply”, which entails improvements regarding user experience. However, adapting the VR technology and content in a way that shows potential to significantly improve user experience is a complex challenge which requires a deeper understanding of a multitude of factors, especially considering that—given the number of different use-cases and stakeholders—there is no *one-size-fits-all* approach to VR hardware and software design.

2.1 Relevant stakeholders

With commercial VR technology still in the early stages of market penetration, cumbersome consumer grade VR solu-

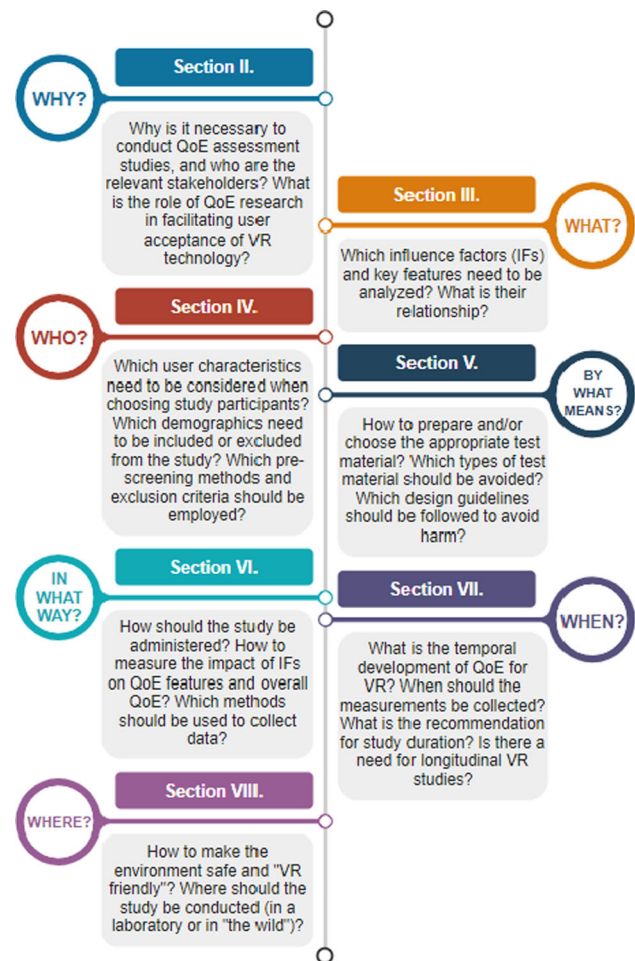


Fig. 1 Paper structure and related research questions

tions leave a lot of room for improvement. Different VR hardware manufacturers are attempting to compete by consistently adding new device features, such as eye-tracking technology (e.g., HTC Vive Pro Eye) or standalone headsets (e.g., Oculus Quest). A diverse selection of different I/O devices (e.g., body tracking technology, haptic devices) has begun to emerge on the market, as companies strive to develop a more natural way to interact with the VE. In addition to efforts invested towards improving single-user solutions, companies focused on gaming and entertainment applications may choose to focus on utilizing VR’s ability to create high levels of co-presence, as evidenced by Horizon Worlds—a social VR world which is in the user testing phase at the time of this writing. As with other multi-user experiences realized through less immersive platforms, social use of VR technology requires adequate network conditions in terms of available bandwidth and low latency. Additionally, a more relevant VR-related challenge for network service

providers pertains to a shift towards split rendering, which utilizes edge cloud infrastructure, as well as the increased use of IoT sensors and actuators contributing toward a more immersive experience, enabled by the capabilities of 5G networks [22].

In the context of VR, the most obvious customer base can be found in gaming enthusiasts seeking a novel, more intense experience. However, VR has long been, and will continue to be, used for various purposes other than its most commonly mentioned use-case—as stated in Perkins Coie [21], aside from gaming and entertainment, immersive technologies (AR and VR) were expected to make a significant impact on the following sectors in the year 2020: healthcare and medical devices sector, education, workforce development and training, manufacturing and automotive industry, marketing and advertising, logistics/transportation, retail/e-commerce, military and defense, commercial and residential real estate, and tourism. The scope of possible VR uses requires thorough research, as each field comes with its own set of requirements in terms of content and input/output devices. However, some aspects and principles of VR design can be generalized across various use-cases and populations. Therefore, there is a need for highly specialized studies using specialized equipment and a target demographic, as well as for more generalized VR QoE/UX studies with a diverse range of participants.

2.2 Understanding user acceptance of virtual reality

To systematise the factors that influence the user to consider using or purchasing VR technology, researchers have developed appropriate technology acceptance models. Sagnier et al. [23] present a VR-adapted extension to the Technology Acceptance Model (TAM) [24]. The model describes the impact of different dimensions of user experience on perceived ease of use and perceived usefulness. Perceived ease of use was found to be significantly influenced by pragmatic quality, i.e., the usability and the utility of the product [25]. Perceived usefulness was found to be significantly influenced by stimulation (a hedonic quality that refers to “the individual’s pursuit of novelty and challenge” [25]) and personal innovativeness. Participants’ intention to use VR appears to be significantly increased by perceived usefulness, and significantly decreased by the severity of cybersickness symptoms, while a significant direct effect of presence has not been found, although it may pose an indirect influence by affecting other variables. A similar TAM-based model, focusing on VR hardware acceptance, is presented by Manis and Choi [26]. The model distinguishes between intention toward using VR hardware, and intention toward purchasing VR hardware. Unlike the model by Sagnier et al. [23], this model does not examine the influence of presence and cybersickness, but it does account for user-related factors such as

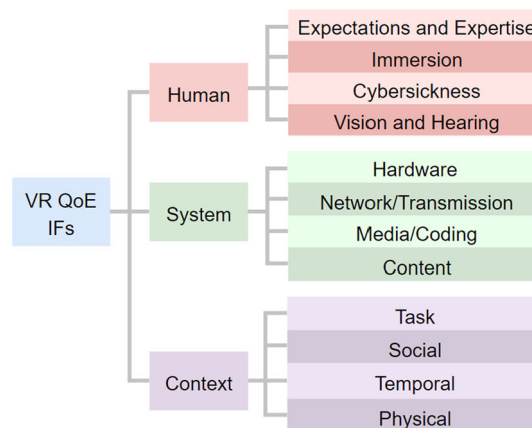


Fig. 2 VR QoE influence factor categories (adapted from [28])

age, previous experience, and the price they were willing to pay for the product. The authors discuss curiosity, perceived usefulness, and perceived ease of use.

According to Whalen et al. [27], QoE in a virtual environment can be maximized by increasing the feeling of enjoyment in users, making it easier for them to accomplish their goals in the context of the application, service, or system, and decreasing discomfort and/or stress. These aspects coincide with the factors presented in Sagnier et al. [23] and Manis and Choi [26], highlighting the connection between VR QoE, intention to use VR software and/or hardware, and the intention to purchase VR software and/or hardware. Therefore, by collecting user evaluations during/after usage of various VR services, realized using different VR systems, researchers gain a deeper insight into multiple variables (often referred to as *influence factors*) affecting user experience, and acquire knowledge regarding their mutual relationship.

3 Quality of experience: influence factors and key features

The Qualinet White Paper [5] defines **influence factors** (IFs) as traits exhibited by the system, service, application, or even users themselves, that may potentially influence QoE of the users of an application or service. Our concise overview of influence factors affecting the interactive VR experience is based on—but not limited to—the classification of influence factors for VR as presented in ITU-T Recomm. G.1035 [28] (Fig. 2).

3.1 Human influence factors

In terms of human (also referred to as user) influencing factors, researchers often choose to examine *dynamic* human factors, such as the current affective state of the user, as well as *static* human factors, which refer to the fixed traits of the participant (e.g., age, sex, etc.). With the common occurrence of VR-related discomfort being an impetus for further research, a high importance is placed on human IFs such as history of illness (e.g., migraine, motion sickness), as well as relevant factors related to vision and hearing. Additionally, previous history of technology use may greatly influence task performance, level of discomfort, and overall satisfaction with the used system. To facilitate comparison of these aspects based on user expertise, participants can be classified according to their general experience with interactive applications (e.g., games) or immersive technology, experience with a particular type/genre of application or, even more specifically, previous experience using a particular application. Considering that VR is still not widely adopted, it should be expected that test subjects may require more time to acclimate to new devices and make more requests for help and instructions [29]. Additionally, when using novel technology, users may perceive their experience as higher in quality due to their own increased interest levels [30]. Further explanations regarding the ways in which previous experience or expectations set by the previous theoretical knowledge of the system/service, may influence QoE, can be found in Sect. 4.1. While listed as influence factors in the ITU-T G.1035 recommendation, cybersickness and immersion may also be examined as QoE features, dependant on other human, system, and context factors. As such, we will describe them in more detail in Sect. 3.4.

3.2 System influence factors

Hardware Influence Factors: Unfortunately, current VR technology is riddled with ergonomic issues. For example, a greater size/weight of VR HMDs may be distracting and uncomfortable to some users and increase the overall physical workload required to interact with the system [31]. As a result of their limitations in terms of adjustability, certain commercial headset designs are not adapted to suit the dimensions of a significant percentage of the population¹. Individuals who use visual correction aids are even more likely to struggle with adjusting the headset to suit their needs [32], especially in case of a system shared by multiple users, as is the case with QoE user studies. Additionally, original versions of contemporary commercial VR headsets have been tethered to the PC and dependent on external sensors, which entails various issues with setup, tracking [33], and

cumbersome cables [34]. However, as of late, standalone versions have been appearing on the market (e.g., Oculus Go, Oculus Quest), offering greater mobility and easier setup at the expense of computing power. VR hardware manufacturers are starting to integrate eye-tracking technology into their HMDs, a feature that can not only be used as an assessment tool (e.g., [35,36]), but also as a tool for optimizing user experience and service performance by enabling foveated rendering [37,38] and fine-tuning user interaction with the VE (e.g., [39]). In general, input (e.g., controllers, gesture control, movement tracking) and output modalities (e.g., headsets, haptic devices) play a significant role in user experience by greatly affecting different quality features. Thus, it is important to pay attention to the possible impact of different device characteristics, such as tracking quality (e.g., [33,40]), latency (e.g., [41–43]), display quality (e.g., [44,45]), and ergonomic design/fit (e.g., [45,46]).

Network Influence Factors: Exploring the impact of networking factors (delay, jitter, bandwidth, packet loss) is currently especially crucial for VR applications centered around 360-degree video streaming (e.g., [47]), although networking issues may also cause significant issues for locally-rendered interactive networked VR applications (e.g., multiplayer games [48], teleoperation [49], or telepresence/collaboration applications). However, 5G and beyond networks are expected to be a disrupting force, revolutionizing the capabilities of immersive interactive VR as we know it. In addition to enabling split rendering, through significant improvements in network bandwidth, latency, and reliability, 5G and beyond networks provide the means for achieving hyper-realistic holographic telepresence. While VR in its current state mostly relies on audio-visual stimuli and body movement tracking to produce a high level of immersion, the significance of haptic technology is expected to increase with the emergence of 5G-enabled Tactile Internet (TI) [50]. *Media/Coding Influence Factors:* This group includes factors related to compression approaches used for encoding audio and video data, as well as other relevant types of information—e.g., point clouds. Aimed at facilitating efficient storage and network transmission, the factors discussed in this paragraph are generally more relevant in the context of 360-degree video (e.g., [51]) and cloud VR (e.g., [52]), compared to synthetic, locally rendered VR services, and are therefore mostly out of scope for this paper. Because of this, we will only briefly touch upon useful sources that may be of interest to readers. For example, Xu et al. [53] present a state-of-the-art overview of 360-degree video and image processing, which includes relevant information regarding perception, quality assessment, and compression methods. With respect to standardization efforts, ITU-T Recomm. P.919 [54] outlines subjective assessment methods for evaluating the QoE of short 360-degree videos. Details are provided on the characteristics of source sequences to be

¹ <https://bit.ly/3eagfFn>.

used, with a wide range of stimuli covering different spatial and temporal complexity, motion, and exploratory properties. Interested readers are further referred to the cross-lab quality assessment tests involving 360-degree videos reported by Gutierrez et al. [55], which were instrumental for the development of ITU-T Recomm. P.919. Among analyzed factors impacting audiovisual quality, the authors consider source content characteristics and uniform and non-uniform coding degradations. In terms of audio, interactive VR services require special consideration of different user movements and positions in relation to other sound sources and listeners positioned within the surrounding virtual space. A paper by Narbutt et al. [56] delves into spatial audio compression and its impact on subjectively-perceived quality. Readers interested in coded representations of immersive media, including not only immersive audio and 360-degree video, but also volumetric data (as discussed in [57,58]), may refer to the ISO/IEC 23090 MPEG-I collection of standards², which contain information on relevant formats, compression methods, quality metrics, implementation guidelines, and reference software.

Content Influence Factors: It is important to take into account different characteristics of the application used in a particular QoE study. In case of interactive applications, such as games, different genres/types can exhibit different levels of sensitivity to different kinds of impairment, such as latency, or produce different levels of immersion and discomfort. Even within the same genre/type, different applications may utilize different mechanics and interaction patterns, realized using different software implementations, which needs to be taken into consideration, as these differences may influence QoE and lead to different conclusions. Notable examples of aspects that are of interest to VR researchers include different characteristics of the avatar (e.g., [59,60]) and the visual environment (e.g., [61]), implementation of the locomotion method (e.g., [62,63]), narrative (e.g., [64,65]), UI design (e.g., [66,67]), etc. With regards to VR gaming, due to similarities with other virtual environments, many relevant content influence factors can be found in the ITU-T Recomm. G.1032 [68] which describes influence factors affecting gaming QoE.

3.3 Context influence factors

Following the discussion of content IFs, different ‘tasks’ performed by end users when evaluating QoE during VR use may be relevant to consider, such as tasks involving different interaction or locomotion techniques. Further, the actual social context is a relevant factor in case of multiplayer/collaboration applications. Arguably, it may be even more relevant for immersive applications compared to con-

ventional platforms. In fact, in addition to an increase in perceived immersion [29,69], VR multiplayer games may result in higher levels of empathy in users when compared to non-VR [69]. User experience may greatly differ depending on the duration and/or frequency of VR use, which impacts the formation of QoE, with the temporal development of QoE, including momentary, reflective, repetitive, and retrospective QoE, explained in further detail in Sect. 7. Physical environment may not be visible to the user immersed in a VE, but environmental variables may be distracting or facilitate the occurrence of cybersickness. Internal and external validity of the results are significantly affected by the setting of the study, i.e., whether it is situated in the field, or in a lab. A more detailed analysis on the impact of the physical context of the study can be found in Sect. 8.

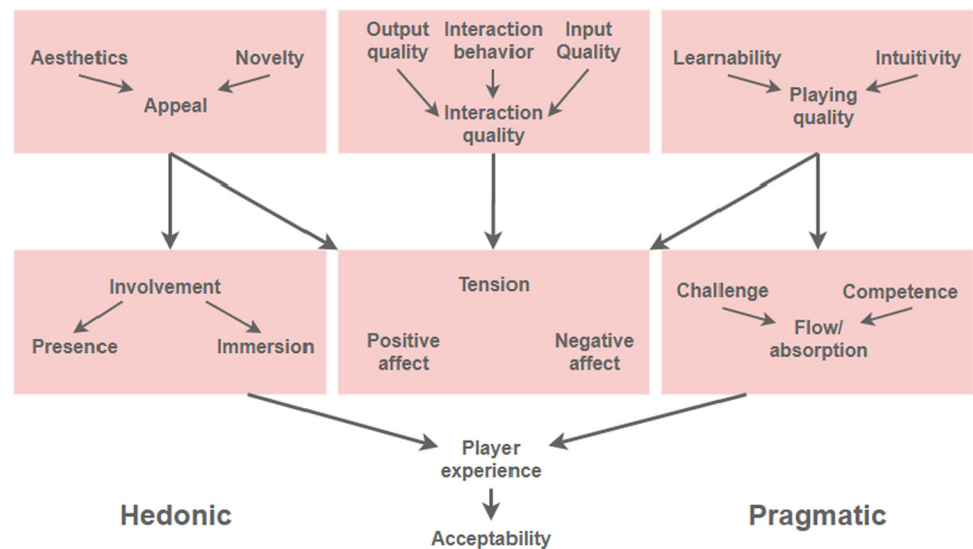
3.4 QoE features

A *quality feature* is defined as “a perceivable, recognized and nameable characteristic of the individual’s experience of a service which contributes to its quality” [72]. Generally speaking, as described in [5,73], quality features can be classified on several levels: level of direct perception (e.g., brightness, contrast, flicker, color perception, loudness, sound localization), level of action (e.g., immersion, perception of space, perception of one’s own movements/motion within that space), level of interaction (e.g., responsiveness, naturalness of interaction), level of the usage instance of the system (e.g., learnability, intuitivity, ease of use, aesthetics), and level of service beyond the particular usage instance (e.g., appeal, usefulness, utility, acceptability). In the context of VR as an interactive, immersive, multi-modal medium, all examples mentioned above can be considered relevant features, but the extent of their individual contributions towards the overall QoE may vary depending on the particular type of VR service.

For example, Fig. 3 displays a taxonomy of gaming QoE features, as presented in ITU-T Recomm. P.809 [70], and based on Möller et al. [71]. However, while certainly transferable to VR, the taxonomy given in Fig. 3 involves some features that may not be relevant to non-gaming interactive VR applications (e.g., tension, challenge). Additionally, it does not include one of the most distinguishing characteristics of the platform—outside of depicted aspects, evaluating VR QoE/UX often includes examining dimensions such as discomfort and cybersickness, which happen as a result of the more physically intrusive nature of the platform, and may significantly degrade user experience. Indeed, aspects such as fatigue and discomfort have previously been recognized as some of the main features of QoE for certain media (i.e., 3D-TV [74]). In line with this, we would like to highlight the need for a general high-level taxonomy (or multiple service-specific taxonomies) of QoE features pertaining specifically

² <https://mpeg.chiariglione.org/standards/mpeg-i>.

Fig. 3 Quality features of gaming QoE (taken from [70]; based on [71])



to interactive VR and incorporating these aspects. Choosing to further focus on features that may be of a particular relevance in the context of VR (in comparison to less immersive media), in the remaining part of this section we present a more in-depth overview of immersion and presence, while an overview of physical symptoms is presented in Sect. 3.5.

3.4.1 Presence, immersion, and related concepts

When discussing user experience related to technologies such as AR and VR, it is important to define *presence* and *immersion*. Schuemie et al. [75] observed that, in literature, the term *presence* generally refers to a self-reported feeling of being transported to a virtual environment (i.e., experiencing a sensation of “being there”). As explained by Slater and Usoh [76], *presence* in the virtual world is the main factor that is specific to VR when compared to different types of media. The authors suggest that presence should be achieved through visual, auditory, tactile, and haptic sensations experienced by the *subject*.

Lee [77] defines presence as “a psychological state in which virtual objects are experienced as actual objects in either sensory or nonsensory ways”, and describes three types of presence. Physical presence refers to a state in which the subject experiences virtual physical objects as if they were actual physical objects. Self-presence refers to a state in which the subject experiences their virtual self (or virtual selves) as if it/they were the actual self. Social presence refers to a state in which the subject experiences virtual social actors (i.e., other humans and/or human-like intelligences) as if they were actual social actors. The definition of social presence encompasses situations that include both one-way and two-way communication, which distinguishes it from the definition of co-presence (i.e., the feeling of being present in a virtual space along with other humans, pertaining to social

interactions with a mutual awareness; [78]), which does not include one-way communication.

With respect to *immersion*, multiple definitions have been proposed in the context of immersive technologies. Witmer and Singer [79] offer a definition of immersion as a psychological state in which a person perceives themselves as being inside of a virtual environment and interacting with it. Slater and Wilbur [80] take a different approach as they describe *immersion* in terms of hardware—more specifically, its ability to provide an experience of artificial reality that can be described as inclusive (referring to the hardware’s ability to block out physical reality), extensive (referring to the extent of independent sensory systems, such as sight, hearing etc., engaged by the hardware), surrounding (referring to the field of view), and vivid (referring to device characteristics such as display quality, resolution, and fidelity).

Cummings and Bailenson [81] performed a meta-analysis based on 83 studies, investigating the relationship between immersion (as a technical quality) and presence. While immersion had a moderate overall effect on presence, certain immersive features (tracking level, field of view, stereoscopy) were found to have a larger impact in comparison to other immersive features, such as image quality, resolution, and sound. These results highlight the importance of spatial cues and self-locating in the presence formation process [82,83], compared to features such as realism and level of detail.

Several distinct terms and concepts are often considered when discussing *presence* and *immersion* (see [84]). For example, Witmer and Singer [79] provide the following definition of *involvement*: “a psychological state experienced as a consequence of focusing one’s energy and attention on a coherent set of stimuli or meaningfully related activities and events”. Weibel et al. [85] define *absorption* as “the capability to concentrate and block out external and distracting stimuli”, and consider it to be one of two indepen-

dent subdimensions of *immersion* (the other being *emotional involvement*).

Slater and Sanchez-Vives [86] use the term *embodiment* to refer to a setup in which the virtual body coincides with the physical body of a user, the user sees the world from the perspective of the virtual body, and there are different types of synchronous multisensory correlation between the two. The visual characteristics of the virtual body (i.e., the avatar) significantly affect the user's experience of the virtual environment. Compared to a generic avatar, embodying a personalized avatar was found to increase the sense of body ownership, as well as perceived presence [87]. Even the user's behaviour, motor functions, and attitude have been shown to change in accordance with the visual characteristics of the corresponding virtual body. The explanations of this phenomenon are given in [88] (the Proteus effect) and [86] (body semantics).

In our brief summary of the aforementioned concepts, we have touched upon certain challenges in terms of wording and nomenclature. For example, although seemingly interchangeable, certain terms (e.g., social presence and co-presence) actually differ from one another in more or less subtle ways, while other terms have multiple (very different) definitions (e.g., immersion). Additionally, due to their relatively abstract, sometimes even vague, definitions, participants may struggle with reporting their subjective perception/evaluation of such features. Researchers are therefore advised to consider these issues when designing a study or comparing other work.

3.5 Physical side-effects

In addition to the previously mentioned QoE influence factors and features, the overall VR QoE is highly dependent on the level of physical discomfort experienced by the user. The common occurrence of physical side effects in participants happens due to the combination of multiple factors, including inherent characteristics of the human perceptual system, static human factors such as age or sex, or technical issues related to application and system design [11].

3.5.1 Cybersickness—definition and symptomology

Immersive technology users commonly experience a state known as cybersickness, which is often likened to motion sickness. Symptoms of motion sickness include emesis (nausea, retching, vomiting), different oculomotor disturbances (e.g., eye-strain, blurred vision) postural instability (also called ataxia) and vertigo [89]. The main distinction between motion sickness and cybersickness is the type of stimulation they tend to be induced by. The main cause of motion sickness is vestibular stimulation [90] (however, visual stimulation may also contribute [91]), while cybersickness can

be provoked by visual stimulation alone. Aukstakalnis [11] provides a comparison between the two in terms of symptomatology. Both motion sickness and cybersickness may cause pallor, nausea, retching/vomiting, increased salivation, increased sweating, dizziness and headaches. In addition to the aforementioned symptoms, Aukstakalnis [11] lists fatigue as a common symptom of motion sickness, and apathy, disorientation, difficulty focusing and blurred vision as common symptoms of cybersickness. However, Mazloumi Gavvani et al. [92] conducted a study comparing symptoms of motion sickness caused by physical movement to symptoms of cybersickness caused by an immersive VR application, and found similarities between symptoms and autonomic changes induced by both types of simulation, leading to the conclusion that motion sickness and cybersickness are clinically identical.

In addition to discussing the differences between “cybersickness” and “motion sickness”, it is important to address the relationship between the terms “cybersickness” and “simulator sickness”. While they are often used interchangeably (or replaced by less common terms such as “virtual reality sickness” or “visually induced motion sickness”, e.g., [11]), and usually examined using the Simulator Sickness Questionnaire (SSQ [93]; see Sect. 6), they differ in terms of context and symptomatology. While the term “simulator sickness” originally refers to the type of discomfort experienced during use of military simulators, cybersickness comes as a result of exposure to VEs. Stanney et al. [94] explain that for simulator sickness, oculomotor symptoms are the most pronounced, followed by nausea and disorientation, while cybersickness results in comparatively higher levels of disorientation, followed by nausea, with oculomotor symptoms being the least prominent symptom group. Additionally, as measured by the SSQ, sickness induced by virtual environment systems results in significantly higher intensity for all three symptom groups compared to simulator sickness [94]. For the sake of consistency, we refer to the state of VR-induced discomfort as “cybersickness” throughout this survey paper, regardless of the exact term used in the cited research.

3.5.2 Factors contributing to cybersickness

Physiological factors LaViola Jr [95] lists three popular theories explaining physiological factors behind cybersickness:

- **Sensory Conflict Theory** states that the main reason behind motion sickness, as well as cybersickness, is the conflict between the vestibular sense and the visual sense [96]; in case of cybersickness this conflict happens when a person perceives movement based on the information on the display, but their body is not actually moving in a way that is suggested by the visual stimulus.

- **Postural Instability Theory** states that cybersickness is caused by an application/service/system forcing the user into a prolonged state of postural instability, meaning that they experience a state of “*uncontrolled movements of the perception and action systems*” which is not adequately minimized [97].
- **Poison Theory** states that cybersickness is caused by an evolutionary mechanism which serves as protection against poisoning; a mismatch between different sensory input systems that happens during immersive application use is incorrectly interpreted by the brain as a symptom of poisoning, which triggers an emetic response in order to empty the stomach of toxic substances [98].

System factors An exploration on the impact of VR hardware maturity on cybersickness is presented in [99]. Specific factors that may contribute to the occurrence of cybersickness are presented below, including factors listed by Aukstakalnis [11], Stanney et al. [15] and LaViola Jr [95]:

- **latency:** the term *latency* refers to the delay that happens between an action performed by the user and the system’s subsequent reaction [100]; latency tends to cause a mismatch between what the user sees, and the proprioceptive sensations the user feels, therefore causing a sensory conflict which may lead to cybersickness.
- **incorrect interpupillary distance settings:** if the lens of the HMD is not properly aligned with the eye, this may trigger the onset of cybersickness symptoms, especially eye-strain and headache [101].
- **optical distortion of scene geometry:** to counteract the phenomenon of pincushion distortion caused by the optical design of the HMD lens, the image needs to be distorted in a way that is equal and opposite to the lens distortion (i.e., barrel distortion); however, this often does not compensate for different eye-lens alignments and subtle changes in eye position, which can lead to issues with depth perception and slight shifts in the perceived position of scene geometry [102].
- **flicker/frame rate:** low frame rate increases the likelihood of flickering, which may cause issues such as eye-strain and nausea [103], although this depends on the user’s individual critical flicker fusion rate threshold [104].
- **position tracking errors:** in addition to standard tracking errors, trackers used in VR systems may produce a jitter effect, i.e., they might move uncontrollably even if the user’s body part remains stationary; this is especially problematic in case of head movements as it shifts the perspective of the user; tracking errors may cause vertigo and difficulty focusing [105].
- **field of view:** a wider field of view, while positively contributing to the sense of presence [106], makes flicker

more noticeable [107] and increases cybersickness [108] due to the sensitivity of the peripheral visual system.

- **scene complexity:** complex environments were shown to produce a significant increase in emetic response [15].
- **implementation of locomotion and camera movement:** vection (perceived self-motion [96]), and especially changes in vection [109], increase cybersickness, while increasing the level of user control over body/camera movement reduces cybersickness [110, 111].

Human factors Various sources (e.g., [11,15,46,95,112–114]) list some of the individual factors that may be linked to a greater susceptibility to cybersickness, such as:

- **age:** cybersickness susceptibility is highest for children between the ages of 2 and 12 [115]; following this early period, it decreases between the ages of 12 and 21 [96] and increases again after age 50 [116].
- **sex:** female users have been found to be more prone to cybersickness [105,113,117].
- **ethnicity/race:** Asian people have been found to be more prone to cybersickness [118].
- **bodily traits and history of illness:** higher body mass index [15], previous experiences with motion sickness and cybersickness [119], migraine propensity [120], etc.
- **behavioral conditions and current state/mood:** inadequate sleep [121], alcohol intake [121,122], acute infections [121,122], being made aware of/thinking about cybersickness [123], strong affective response to stimuli [114].
- **psychological traits and personality type:** neuroticism [114], anxiety [124], low self-efficacy towards technology [113], low perceived sense of direction [113], lower preference towards adrenaline sports [99].

3.5.3 Adaptation

Wang and Suh [125] discuss different types of adaptation mechanisms (behavioral, cognitive and physiological adaptation) used for counteracting cybersickness. When users experience cybersickness, they tend to perform certain actions as a way to mitigate their symptoms, which is referred to as *behavioral adaptation*. For example, these actions may include taking a break, moving in a different way or adjusting the headset. *Cognitive adaptation* refers to the user’s choice to withstand the symptoms of cybersickness because they consider them to be a normal part of the experience. With continued and repeated use, users seem to acquire a certain level of resistance to cybersickness—this type of adaptation is referred to as *physiological adaptation*.

3.6 Digital eye strain and ergonomics

While the issue of cybersickness has already been researched and discussed in a large body of work, other types of VR-related discomfort have not yet garnered a lot of attention [126]. This gap in research is examined in a recent paper by Hirzle et al. [127]. When comparing the relevance of three symptom categories—referred as simulator sickness, digital eye strain, and ergonomics—in an online study conducted on 352 frequent VR users, the authors found that the majority of participants considered simulator sickness to be less relevant compared to both remaining categories.

Focusing on ocular symptoms, there are multiple potential causes for discomfort related to head-mounted displays, such as delay, flicker, resolution, image motion and binocular imperfections, which may be caused by optics (e.g., image blur, shift, rotation), the use of filters (e.g., luminance, color, contrast) or stereoscopic disparity [128]. As explained in [129], visual fatigue in VR is mostly a result of vergence-accommodation conflict (VAC) [130,131], which occurs in case of a mismatch between the accommodation distance and the rendered image distance, but may also happen because of motion (especially vibrational motion [132], which requires the gaze to be stabilised by the vestibulo-ocular reflex—VOR).

With regard to other ergonomic issues of VR systems, we highlight the following examples:

- **heat:** heat development on the VR HMD may cause significant discomfort [45] and increase sweating [127]; the uncomfortable feeling of increased body heat may also come as a result of physical activity required by some types of VR applications;
- **weight:** as a result of wearing an HMD, VR users tend to modify their posture, thus stressing the musculoskeletal system (especially head and neck areas [133]); aside from the total weight of the HMD, another aspect that affects comfort is its distribution, as an imbalanced HMD design places greater torque around the neck of the user [31,45];
- **pain and muscle fatigue:** the aforementioned neck strain that comes as a result of HMD weight can become exacerbated by frequent head turns as the user looks around the virtual environment, while frequent mid-air arm movements required by certain application mechanics are likely to produce a feeling of heaviness and fatigue in the arms and shoulders, an effect referred to as *gorilla arms* [134]; when using VR applications that require users to squat or bend down, users may experience fatigue and/or pain in different areas of the back and legs; the issue of muscle strain is especially relevant for exergaming solutions, especially if they require the use of additional exercise equipment;

- **adjustability issues:** commercial HMD designs incorporate wheels and straps for a more precise adjustment of headset fit and interpupillary distance; however, due to the headset weight and other design factors, VR users often struggle with finding the right balance between too tight and too loose, both of which can be uncomfortable for the skin or different regions of the face and head [46]. For example, based on our experience in conducting VR studies involving numerous users using various commercially available HMDs, in cases where the HMD is not properly fastened, users need to take frequent pauses for readjusting the setup, thus possibly breaking the immersion, while excessive tightness or friction between the HMD and the skin may leave the user with lingering facial redness and headset lines; the problems associated with headset adjustability are likely to be even more pronounced for users with refractive errors, especially in case they need to use their prescription eye wear underneath the headset [46]; HMD owners may find a solution in purchasing prescription lens inserts, but the issue remains for researchers conducting studies on multiple subjects using the same HMD.

3.7 Cognitive effects

In addition to potentially triggering different types of discomfort, the use of VR may affect certain cognitive processes. The impact of VR use on reaction time, mental rotational activity, perceptual speed and visual working memory was examined in a paper by Mittelstaedt et al. [135]. Immersion in VR did not result in declined performance for perceptual speed and visual working memory, and even produced an improvement in the processing speed for mental rotation. However, the authors report an increase in reaction time after VR exposure—an effect that has also been observed in [136,137], although its etiology is not yet fully understood. While results presented in [136,137] provide a link between the slowing down of reaction time and cybersickness, Mittelstaedt et al. [135] presented other possible explanations including visuomotor adaptation due to sensory mismatch and temporal adaptation to the slight delays introduced by the used devices. Szpak et al. [138] observed the impact of VR exergaming on decision time and movement speed during a multiple choice reaction time test. The authors did not find a significant effect for decision time, while motor movement time improved after 10 min of VR exposure, although the effect did not linger for long. This improvement in reaction time may be explained as a result of the physical activity required by the game.

While it is unclear whether the effect of VR on reaction time is significant enough to raise concern regarding the dangers of operating cars and heavy machinery immediately post-VR exposure, it is important to note that the average

increase remains below 50 ms across different studies. Additionally, this effect may only be short-lived, and therefore easily mitigated by incorporating a short (e.g., 40 min [138]) wait period before attempting to perform any potentially hazardous activities. However, the value of measuring cognitive performance in relation to VR use extends beyond the implications regarding user safety—e.g., cognitive performance measures may provide a better understanding of cognitive processes required for adequate functioning in an artificial environment, or serve as a benchmark for assessing cognitive fatigue and quality/naturalness of different interaction mechanics and input devices.

4 The importance of pre-screening and participant choice

Gravetter and Forzano [139] state that external validity of a user study in the field of behavioral sciences refers to “*the extent to which we can generalize the results of a research study to people, settings, times, measures and characteristics other than those used in that study*”. Therefore, when performing QoE assessment, researchers should aim to formulate their research methodology in a way that would enable the study to be reproduced with similar results. Additionally, to achieve valid results, study conditions should be adapted to resemble a real-world usage scenario. An important component to achieving a high level of external validity is the extent to which we can generalize the results of a study from a sample population to the general public. Unfortunately, with regards to VR, this proves to be a significant challenge.

4.1 Experience and preconceptions

The so-called *halo effect* [140,141] happens when judgments about unknown characteristics of an entity (e.g., person, object, system) are made based on its evident and/or previously known characteristics. At the time of writing, VR is still widely considered to be a *niche* type of technology, owned by a small number of early adopters (e.g., gaming enthusiasts). Due to the relative scarcity of VR systems, and their reputation for creating a more immersive experience in comparison to virtually any other platform, the possibility of participating in a VR research study tends to arouse interest in potential subjects, especially those of a younger demographic. This preconceived enthusiasm might influence study results by clouding the participants’ judgement of the system or the application.

On a related note, the *novelty effect* happens when test subjects’ perception and responses in a research study setting (which is considered to be a novel situation) deviate from their perception and responses in a real-world situa-

tion [139]. With VR, the users may also be influenced by the perceived novelty of the platform itself. While Fairchild et al. [142] noted that novice users may experience VR in a negative way, Hupont et al. [29] attribute positive affective states experienced by test subjects to the novelty of the VR platform. Whether positive or negative, the potential impact of platform novelty on user experience should be considered when choosing test subjects and/or interpreting study results.

Considering study results serve as input for creating QoE models and guidelines to be used by hardware manufacturers, content developers, and network service providers, we believe that researchers should try to avoid relying on conventional convenience sampling, which is likely to result in a very inexperienced sample of participants due to the aforementioned issue of relative scarcity of VR systems. Instead, assuming the necessity of nonprobability sampling approaches, researchers could lean towards quota sampling as a way to achieve a more balanced distribution of users with different levels of experience. If advanced users are not available for study participation, researchers could look into participant recruitment via crowdsourcing platforms [143]. Alternatively, some of the less experienced subjects may be given several training sessions prior to the actual test session as a way to mitigate the impact of aforementioned psychological effects (e.g., cybersickness adaptation training, as discussed in [144]).

4.2 Ethics, health and safety

Madary and Metzinger [145] highlight the importance of pre-screening as a way to remain in compliance with the principle of *non-maleficence*³, which instructs researchers to construct their experiments in a way that ensures no significant or long-term harm would come to subjects as a consequence of participating in the study. The authors especially warn about the well-being of participants with psychiatric disorders (whether diagnosed or undiagnosed). It is important to stress, though, that VR is often used in treatment of certain psychiatric disorders, in which case pre-screening should be employed with the purpose of finding participants with that particular disorder. However, even in those cases, researchers should remain mindful of the possible psychological impact, and exclude participants whose psychological vulnerabilities or other conditions put them at risk. Therefore, depending on the aim of the study, researchers should define appropriate *exclusion criteria* by using specialized questionnaires to assess whether the user has previously exhibited or currently exhibits signs and symptoms (e.g., dissociative experiences, psychotic episodes, suicidal ideation) of certain disorders that may get aggravated by the experience. Behr et al. [146] suggest screening participants for space-related

³ <https://www.apa.org/ethics/code/>.

phobias (e.g., claustrophobia, agoraphobia), as well as other phobias specifically related to the test material.

Lewis and Griffin [147] offer suggestions for screening participants prior to the clinical use of VR. They advise against including participants who are ill with diseases such as influenza, ear infections or ocular defects, suffering from balance disorders and/or taking medication that affects visual or vestibular function, currently under the influence of alcohol, or prone to motion sickness or cybersickness. These pre-screening guidelines may also be utilized for non-clinical studies involving VR exposure. In general, it is often advised that people who show high levels of sensitivity to cybersickness should not be exposed to VR [146] even in a research setting. However, from the perspective of product developers, including more vulnerable participants allows for a deeper level of insight, which can then be utilized to improve the application or system. Therefore, an alternative approach [148] is to use questionnaires to purposefully select subjects who have previously showed signs of cybersickness or motion sickness, as well as inexperienced and elderly users, who might be more sensitive, while also choosing to invite a larger number of participants in case sensitive users have to terminate the experiment due to the onset of symptoms. In terms of visual impairment and ocular symptoms, participants may be excluded based on their scores on visual acuity, color-blindness, or stereopsis tests.

Researchers exploring interactive VR may benefit from examples and guidelines regarding the inclusion of users with disabilities in gaming user research, presented in [148]. For example, study administrators should be on the lookout for situations that might make participants feel frustrated and vulnerable, such as not being able to successfully perform the required activity, being tested in a group setting or involved in a multi-player game. In order to adapt the process to the specific needs of each participant, researchers may need to consult with medical experts, therapists, and/or caretakers, as well as with individual participants if necessary. In general, it is very important to keep the whole process of testing as flexible and adaptable as possible.

4.3 Diversifying the study population

According to research, sex plays a significant role in evaluation of vitally important elements of the VR experience, i.e., the perception of presence [149] and susceptibility to cybersickness [117], with researchers suggesting that VR technology tends to be more adapted to male users. However, to fully understand these implications, further comparisons need to be drawn based on experimental data. To be able to compare sex differences, scientists should strive towards achieving a balanced sex distribution of study participants, while also considering gender differences in experience with VEs, such as games.

VR systems and applications are mostly geared towards a younger, tech-savvy, audience. On top of that, college students are over-represented in user studies regarding human psychology and behaviour [150]. While quantifying the percentage of college students participating in VR user studies is beyond the scope of this paper, we are under the impression that recruiting this demographic is a common practice among VR researchers. Unfortunately, this also means that their findings are not necessarily applicable to a wider range of VR users, as differences in VR user experience between users of various age/age groups have already been noted by researchers (e.g., [151,152]). For example, the perceived ease of use with VR technology [26] may differ based on age/age group, and age may play a certain role in the susceptibility of cybersickness (as mentioned in Sect. 3), illusion of body ownership [153], as well as immersion and presence [154]. The oversampling of the young adult demographic can be prevented/counteracted by incorporating participants of different ages into the study population, or designing studies specifically for underrepresented age groups, such as children, or the elderly. However, incorporating these age groups may require special consideration.

Researchers (as well as VR system manufacturers and VR application developers) warn about the unknown impact of VR use on children and young adults with regards to their psychological and neurophysiological development [145]. However, Tychsen and Foeller [155] conducted a user study on 50 children (aged between 3 and 10 years old) and reported that 94% of participants experienced no significant differences in postural stability, as well as no observable symptoms of dizziness and cybersickness, while playing a flying VR game. Additionally, measuring horizontal VOR in a small subset of participants (5 children)—before and after VR use—yielded no evidence of vestibulo-ocular maladaptation. However, subjective scores for cybersickness, dizziness, eye strain, and head/neck discomfort post-VR were higher compared to baseline, although the authors note that the observed difference is only statistically, but not biologically, significant. While there is a need for further research (especially longitudinal studies) in this area, according to this study, limited VR exposure (such as participating in a short study session) is not likely to cause long-term psychological or physiological damage to young participants if appropriate precautions are taken (e.g., limiting the session duration, repeatedly reminding participants to terminate the experiment if they experience fear or discomfort, exposing them to age-appropriate material only).

In terms of elderly users, VR may be used for entertainment, to diagnose and treat conditions such as Alzheimer's disease, or as an aid in physical therapy, helping users to improve their balance or motor skills. However, when designing studies for this demographic, it is important to adjust the content and/or study methodology to their specific limita-

tions regarding mobility, cognitive abilities, and computer literacy. As with other participants with disabilities, elderly users may not be able to successfully participate in studies that are not designed for them specifically, as they may find it difficult to navigate the application or perform certain physical movements (e.g., turning [156]), which can lead to frustration and lowered confidence [148]. Elderly users may experience difficulty with vision and hearing [156], which can be counteracted by adjusting the volume of the test material, presenting the instructions in a clear, comprehensive way, and repeating them as needed. Furthermore, it is important to keep in mind that older individuals may be at a higher risk of falling during VR experiences [157].

5 Guidelines for preparation of appropriate test material

The test material used for conducting VR user studies depends on the aim of the study, and can range from applications with a practical purpose, such as those intended for therapeutic use (e.g., physical therapy, cognitive therapy, phobia treatment), educational applications or scientific visualisations (e.g., medical applications, military training), to applications intended for entertainment purposes (e.g., games, drawing in VR). Test material can be developed specifically for the study, or it may be a short sample of an existing application. The latter option is especially appropriate for VR gaming studies. As suggested by the ITU-T Recomm. P.809 [70], researchers should carefully select a sample that displays a mechanic that is typical for the game (or another type of application). If using a fixed level of difficulty, researchers should aim to select a sample that is appropriate for participants with various levels of experience. Otherwise, they may choose to keep it adjustable, so that it can be adapted to fit the skill level of each participant. Prior to conducting the actual study, test material should be thoroughly examined to ensure that the application runs smoothly. As discussed in [148], the frustration caused by encountering bugs and crashes during a test session is likely to degrade reported user experience.

Schatz et al. [158] highlighted the deficit of standardized VR content as well as a lack of standardized test tasks that would enable the reproduction of user studies across different laboratories and research groups. An example of such a test task can be found in [33], where the authors use a simple pick-and-place task to compare the performance of different VR systems. While design, development, and distribution of standardized test content remains an open challenge, researchers can facilitate comparison between studies by describing the used application, as well as chosen methods of interaction and locomotion.

5.1 Ethics, health and safety related to choice of test material

As previously discussed, researchers are expected to follow the principle of *non-maleficence*. Thus, the material chosen or created for the study should not inflict significant or long-term psychological or physical harm.

5.1.1 Avoiding psychological harm

Virtual environments (VEs) differ from other types of media based on two main characteristics [159]: *saliency* (i.e., VEs provide a more salient/vivid experience by combining multiple sensory stimuli) and *agency* (i.e., VEs allow the user to interact with their surroundings). The *information overload* during VR use [146] is a result of high levels of *saliency*, enabled by the inherent multimodality of VR systems which expose the user to various sensory stimuli (predominantly audio-visual, often haptic) at the same time, combined with the system's intrusiveness. Unlike hand-held or desktop displays, VR HMDs are strapped onto the user's head and often equipped with integrated headphones or used with external earphones. This setup is purposefully designed to “*override*” any audio-visual input from the real world, leading to greater immersion, but making it difficult for users to avoid or escape [146] the artificial sensations they find uncomfortable or overwhelming.

The small lens-to-eye distance in HMD-based VR systems may cause the user to experience the virtual world more concretely compared to other platforms, yet even in CAVE-based VR studies, participants have been shown to respond to a stressful situation with subjective, behavioral and physiological reactions, despite being aware of the artificiality of the presented stimuli [160]. Segovia et al. [161] use HMD-based VR to demonstrate the impact of situations experienced in immersive VR on the moral identity of the user. This connects back to *agency* as a defining aspect of virtual environments. *Agent regret* refers to the phenomenon of a person experiencing more guilt after performing an *innocent* action that led to a negative outcome of a certain situation, than they would have if they merely witnessed the negative outcome without having performed any action at all [159,162]. A VR application which contains disturbing material may therefore interfere with the user in a more significant way compared to e.g., watching a video based around a similar theme.

Spending a longer period of time in VR may cause issues with discerning between the virtual world and the physical reality, as seen in [163]. While short-term effects, such as experiencing so-called *Game Transfer Phenomena* (GTP) [164] shortly after exposure to a non-stressful VR application, may not pose a significant threat to psychological and emotional well-being of the participant, the impact of immersion may be increased or prolonged in case of exposure

to stressful, scary or otherwise disturbing content. Despite obviously not being real, disturbing media content (e.g., a horror movie) can leave a long-lasting, even lifelong, negative impact on the consumer, resulting in *media-induced trauma* [165]. However, a study done by Lin [166] showed that lingering effects of a horror game in VR may not be as common or as intense as one might expect, considering only a small number of participants reported experiencing them the day after the study. Despite these findings, it is advisable to avoid exposing users to uncomfortable content unless it is highly relevant for the specific study. In case the test application involves potentially unnerving material, participants should be warned in advance, as well as encouraged to pause or terminate the experiment by taking off the HMD. The test application should include a virtual safe space [167] which allows participants to immediately (i.e., with a button press) escape the anxiety-provoking stimulus without physically removing the VR equipment, as it might be difficult to loosen the straps and take off the headset quickly whilst holding the controllers. Aside from the ethical issue of being exposed to potentially traumatising content, witnessing disturbing events in VR is likely to produce physiological reactions which, if registered by devices such as an EEG or a heart-rate monitor, may complicate subsequent analysis by increasing the ambiguity of the results.

5.1.2 Avoiding discomfort and cybersickness

Considering that, even after decades of ongoing research and development, cybersickness in VR still remains a pressing issue for VR scientists, developers, and users alike, there is certainly a need for further research in this area. However, in order to avoid inflicting physical harm while researching the condition or conducting VR user studies in general, researchers should choose or develop test material based on the state-of-the-art knowledge of design factors that might impact the occurrence of cybersickness and other types of discomfort. A compilation of guidelines and useful findings is presented in Table 1.

6 QoE assessment study methodology

In accordance with the principle of respect for persons [187], the autonomy of each study participant has to be respected, which means that researchers have the responsibility to provide relevant information about the study and ask for consent prior to actual data collection. After the consent form is signed, a pre-test questionnaire is given as a way to collect personal information about the participant. Similarly to questionnaires used in gaming research [70], pre-test questionnaires used in VR studies usually encompass questions about the basic demographic data (age, sex/gender, profes-

sion, ethnicity), as well as inquiries about the skill level and prior experience. Participants may be inquired about their history of illness, or asked to fill out specialized questionnaires as a way to assess their personality traits, or psychological and/or physiological sensitivities. Along with questionnaires, researchers may choose to include standardized vision acuity tests in their pre-testing process. Acquired information aids in later analysis and interpretation of study results, but it can also serve as exclusion criteria. Participants should be made aware that they are allowed to pause or terminate the experiment at any time. Instructions regarding equipment, material, and assessment methods should be carefully worded, easy to understand, and presented to each participant in an identical way, which helps mitigate instruction bias [148]. Following the instruction phase, participants are equipped with VR and measurement devices, the positioning of which may require some assistance from the administrator. It is highly advisable to sanitize the equipment (headset, handheld controllers, and any other devices that come into contact with the participant) before each session, which is especially relevant in light of the recent COVID-19 pandemic. If possible, study administrators should provide each user with a disposable mask that provides a barrier between their skin and the headset. It is advisable to warn participants against operating a vehicle following the exposure to VR content. Although there are no official guidelines at the time of this writing (to the best of our knowledge), the duration of the recommended waiting period will likely depend on the intensity of the application and the duration of the VR exposure [95]: several minutes of exposure to a commercial VR application may require only a short 30–45 min waiting period, while a longer exposure to a flying simulator may require a waiting period of 12 to 24 h. The last step before the actual testing phase is a short tutorial session which facilitates adaptation to the application and the technology. Details regarding temporal or environmental aspects of study design are discussed in Sects. 7 and 8, while the remaining part of this section provides an overview of commonly used assessment methods in VR user research.

6.1 QoE assessment methods

At the time of this writing, there is no standardised methodology for assessing the QoE of VR applications (although efforts are underway in the scope of ITU-T Study Group 12 [188]). However, there are a number of instruments that have been used across various studies addressing the assessment of QoE-related features, such as immersion and presence, as well as side-effects such as cybersickness.

6.1.1 Subjective methods

The use of questionnaires is the most common subjective method used in QoE studies, although it may be supple-

Table 1 VR application design guidelines and findings for mitigating cybersickness, discomfort, and other health risks

Category	Guideline	Ref.
Temporal aspects	Framerate should be maintained at 90 fps or above.	[148]
	Motion-to-photon latency should remain constant at 20 ms or below.	[148,168]
Camera	Camera movement should directly correspond to HMD movement.	[148]
	User should stay in control of camera movement at all times.	[46,111,148,169]
	Incorporating the zooming option may worsen cybersickness.	[148]
Movement/locomotion	Lifting the camera view high above ground may increase vertigo, while lowering it close to the ground may increase cybersickness.	[148,168]
	Completely omitting movement/locomotion from the application is the most effective way of preventing discomfort.	[168,170]
	Rolling motion/rotational optic flow may worsen cybersickness.	[148,171,172]
	Acceleration (including deacceleration) is the primary cause of discomfort in VR.	[109,111,148,168]
	Blinks and snap turns (up to 8 times per second) disrupt the perception of motion, decreasing discomfort.	[168,173]
	Discrete motion (i.e., teleportation) minimizes cybersickness, while continuous motion exacerbates it.	[148,168]
	Avoiding visual motion cues in the peripheral vision, by vignetting/tunnel vision or putting the user in a virtual enclosed vehicle during continuous movement, may reduce discomfort.	[111,148,168,174]
	Emulating experiences that are likely to cause motion sickness in a real environment (e.g. a rollercoaster ride) is likely to induce cybersickness in a virtual environment.	[148,175]
	Head bobbing and moving across uneven terrain (e.g., staircase, hills) can trigger cybersickness.	[46,176]
	Forward movement is more natural and comfortable compared to lateral movement.	[168]
Allowing the user to stay in control of their movement minimizes cybersickness.	[46,111,168,177]	
Moving through confined spaces is less comfortable compared to moving through large, open spaces, and moving parallel to a nearby wall is less comfortable than moving parallel to a distant wall.	[168,178]	
A small input delay (70–100 ms) between activating the control for movement and actual beginning of movement reduces discomfort by allowing the user to anticipate/ predict the sensation of motion.	[168]	

Table 1 continued

Category	Guideline	Ref.
Embodiment	The position and orientation of virtual hands needs to be aligned with the hands of the user to avoid discomfort caused by proprioceptive mismatch.	[168]
	Hand animations that are in conflict with the user's real movement should be avoided.	[168,179]
	Omitting virtual representations of arms above the wrist may decrease proprioceptive mismatch that happens in case of incorrect or imprecise elbow/ shoulder tracking.	[168,180]
	If using a full body avatar, its movements should remain aligned with user movements to avoid discomfort caused by proprioceptive mismatch.	[168,180,181]
	Discomfort may also be reduced by incorporating an avatar that prepares the user by foreshadowing impending movement.	[111,168,182]
	Motion blur, blurriness, flickering and flashing lights/elements should be avoided.	[111,148,168]
	Using darker colors is preferable, as brighter colors are more likely to produce display flicker, which is uncomfortable for users, as well as contribute towards increased visual fatigue.	[168,183]
	Visual effects (e.g., particle effects, bloom) should be rendered in a way that can be seen with both eyes with regards to binocular disparity.	[168]
	Repeated patterns and textures, and optical illusions of depth may lead to misperception of depth, which may be uncomfortable for the user.	[168]
	High-spatial-frequency textures are a potential trigger for cybersickness and photosensitive seizures and should therefore be avoided.	[168,184]
Focus	Higher levels of detail and realism in visual design increase cybersickness.	[175,184,185]
	Optical distortion introduced by the HMD lenses should be counteracted by applying appropriate distortion correction (by post-processing in the SDK).	[168]
	Incorporating dynamic focus may decrease feelings of discomfort (users should be able to focus on objects regardless of their distance or position in the FOV).	[111]
UI	Transition focus speed in applications incorporating dynamic focus should be kept at a minimum of 500 milliseconds.	[111,186]
	Positioning UI elements in a manner that requires large and sudden head and/or eye movements from the user should be avoided.	[148]
	UI elements should be incorporated into the scene, since improperly positioned/ rendered elements may occlude the scene in an unnatural way and lead to the misperception of depth cues.	[168]

mented with other methods, such as interviews and diary entries. In most cases, individuals are asked to fill out questionnaires directly related to tested scenarios either during or immediately after testing. Most commonly, users are required to mark their answer on a rating scale. Users may be asked to provide their rating of the overall QoE or its individual dimensions. Instead of using individual questions, researchers often choose to use more established multi-item questionnaires designed to measure a certain aspect (or multiple aspects) of quality. For example, usability can be evaluated using the System Usability Questionnaire (SUS) [189], while the Self-Assessment Manikin (SAM) [190] may be used to assess the user's affective response.

Certain questionnaires used in QoE research cover a diverse range of features and are intended to be used as a single tool for the evaluation of the overall quality, such as the Game Experience Questionnaire (GEQ) [191,192] or the Player Experience Inventory (PXI) [193,194] which are designed for the gaming use-case. Unfortunately, due to the specific characteristics of interactive VR, questionnaires that were initially developed with non-immersive platforms in mind can not be used on their own (i.e., they need to be combined with other measures, which can sometimes be fatiguing for participants and complicates subsequent analysis of results) as they do not include certain aspects that are especially relevant to the VR platform, such as discomfort and cybersickness. This highlights the importance of developing questionnaires that can be used for the evaluation of QoE/UX based on specific features that are relevant for interactive VR. An example of a VR questionnaire that evaluates multiple different features (i.e., general user experience, game mechanics, in-game assistance, symptoms and effects induced by VR) is the Virtual Reality Neuroscience Questionnaire (VRNQ) [195], but its use is limited to VR gaming, rather than VR in general. Tcha-Tokey et al. [196] developed a more general-use VR UX questionnaire comprised of nine subscales: presence, engagement, immersion, flow, emotion, skill, judgement, experience consequence (which measures symptoms of fatigue and cybersickness), and technology adoption.

The problem with subjective measures is that they are self-reported and therefore cognitively mediated, which leads to distortions and undermines their validity. E.g., participants often tend to avoid either extreme of the scale (*central tendency bias*), or respond in an excessively positive/agreeable manner (*acquiescence bias*), while further issues stem from the improper or unclear wording of questions themselves. Lastly, since participants' view of the real world is obscured by the VR HMD, their answers are often noted by an administrator, which may influence the participant [197]. Therefore, if possible, subjective assessment questionnaires should be integrated into VEs used for testing [198].

6.1.2 Objective methods

In addition to subjective methods, objective methods (physiological, behavioral, and task performance measures) are often used to assess user experience in a less biased way. Physiological methods are based on measuring different physiological signals such as electrocardiography (ECG), electroencephalography (EEG) and galvanic skin response (GSR). Due to their design, certain medical instruments used to collect this data may hinder user experience and degrade QoE scores, so less intrusive devices, such as fitness bands and smart watches, can also be used for collecting physiological signals [199]. As discussed in Skorin-Kapov et al. [200], the use of psychophysiological measurements in assessing user experience improves existing QoE models, especially in terms of user-related factors, and mitigates issues stemming from the use of self-reported assessments [201,202]. However, it should be noted that it can be challenging to adequately recognize the affective state of the user based on physiological measures only, as different states may be indicated by very similar physiological symptoms [203,204]—for example, both excitement and stress tend to increase the heart rate of the user. Furthermore, certain methods for measuring physiological signals appear to be sensitive to noise introduced by head movement (e.g., EEG [205]), while others, such as functional magnetic resonance imaging (fMRI), require complete stillness. Therefore, the results of such methods may not be accurate unless the study happens to be consciously designed in a way that aims to keep the user as stationary as possible. Since head movement in VR is not only extremely common, but also highly encouraged through VR application design, the degree to which the results acquired in stationary conditions can be considered representative of realistic VR use is yet to be determined.

Behavioral methods refer to methods that are based on observing and tracking user behaviors, such as physical movement (e.g., “ducking” to dodge an approaching virtual object [206]) and social interaction (e.g., moving away from an avatar or an embodied agent [207]). To assess user preferences or adaptation mechanisms in the context of VR application use, researchers may decide to track and categorize different actions that the user chooses to perform inside of the interactive VE. In addition to larger bodily movements and conscious actions, researchers may choose to observe more subtle behaviours by incorporating methods such as gaze tracking and emotion recognition, made possible by the growing inclusion of eye tracking and facial recognition technology in more recent headsets.

In general, user performance in multimodal interactive systems, such as VR, encompasses three components [208]: perceptual effort, cognitive workload, and physical response effort. Task performance measures (e.g., time to complete task, measures pertaining to spatial and temporal accuracy)

aid in quantifying the effort produced to accomplish a task, and may serve as an objective indicator of the impact of different factors on the users ability to interact with the service in a successful and efficient way, thus providing an objective measure for the evaluation of QoE features such as ease of use and interaction quality. However, to increase the chances of obtaining conclusive and valid results, it is important to choose tasks and measures that are relevant to the observed system/environment.

6.1.3 Measuring presence and immersion

Subjective measures: Subjective ratings are commonly collected using questionnaires, with a concise list of questionnaires addressing *presence* and *immersion* presented in Table 2. For a more comprehensive analysis of presence questionnaires, the reader may refer to [75,84], while an overview of studies related to presence and immersion, including information regarding presence questionnaires, is presented in [81]. While the majority of immersion/presence questionnaires are constructed from multiple items, Jennett et al. [219] report that a single-item questionnaire (i.e., “rate how immersed you felt from 1 to 10”) appears to be a reliable measure of immersion as well, which has recently been supported by findings presented in [223].

However, the practice of using questionnaires as a primary method of measuring presence/immersion has been heavily criticised for various reasons. For example, such abstract constructs tend to be loosely defined and therefore open to interpretation, as discussed in [224]. Furthermore, asking users to report their sense of presence/immersion in the middle of the experience that is being evaluated will likely lead to its disruption (as discussed in [225]), and reporting the sense of presence/immersion after the experience has ended relies on potentially inaccurate recollection [226]. Keeping within the context of subjective, self-reported measures, instead of presenting questionnaires during or after a session of VR use, Slater and Steed [225] suggest tracking breaks in presence (BIPs) during exposure to a virtual environment. This method requires users to report transitions from the state of absorption in the virtual environment to the state of being “back to reality”.

Objective measures: Notable behavioural measures for assessing presence are reactions to conflicts between virtual cues and real cues [227,228] and actions such as reflex responses to virtual events [206]. On a related note, Lepecq et al. [229] propose *afforded actions* as a way to evaluate presence in VR environments, as users tend to perform behavioral transitions (body rotation) to adapt to the characteristics of the presented virtual environment (narrow virtual aperture) in relation to their own body characteristics (width of shoulders). This is similar to the approach taken by Usoh et al. [216], who observed the path taken by users immersed in

a virtual environment to see whether they would choose to step over an unsettling virtual pit or try to walk along its edge as they traverse across the room. Moreover, the aforementioned paper by Usoh et al. describes a combination of different measures (behavioral as well as subjective) that are used to construct the measure of *behavioral presence*—the degree in which “actual behaviors or internal states and perceptions” suggest a sense of being in the virtual environment instead of the real, physical one.

Given that VR experiences may be able to elicit reactions and emotions that are comparable to those that arise in real-world situations [75], immersive technology has a wide field of application in social science research. Scientists often use methods that rely on observing and tracking human interaction in multi-user environments, and are therefore used for the exploration of constructs such as social presence and co-presence. For example, as presented in [230], task performance metrics can be used as a way to measure social inhibition and facilitation (e.g., [231,232]) when faced with real or virtual humans, measuring interpersonal distance and personal space (e.g., [233,234]) is used in the context of proxemics research, and tracking eye gaze and facial expressions can provide information regarding the affective state of the user in a social situation (e.g., [235,236]).

Presence can also be assessed by examining physiological measures (e.g., GSR, EEG, heart rate, body temperature). Meehan et al. [226] listed subject bias and inaccurate recollection as disadvantages of subjective measures, while also taking note of experimenter bias that may occur with the use of behavioral measures. Thus, the authors reported looking for a measure of presence that meets several criteria: validity in terms of correlation with broadly accepted subjective measures of presence, objectivity, sensitivity to different levels of presence, and reliability/repeatability. Comparing different physiological measures (heart rate, GSR, body temperature), heart rate was found to be the best in meeting the abovementioned criteria, followed by GSR [226,237]. As opposed to heart rate, GSR did not show promise as a between-user measure.

Bouchard et al. [204] criticise the use of physiological measures of presence, as changes in heart rate and GSR are well-established measures of anxiety, and therefore likely indicate anxiety—rather than presence—in stressful virtual environments. Thus, the authors describe these measures as “at best, proxy measures of presence in anxiety-related contexts” (an example of this can be seen in [61]). Bouchard et al. also argue that exposing the user to both real and virtual situations to see whether they produce similar physiological responses is a much better way of measuring presence than the common approach of measuring changes in physiological signals during various virtual scenarios. An example of an approach based on comparing physiological signals in a virtual situation and a real situation can be seen in [238], as

Table 2 Overview of presence questionnaires (adapted from [75,84])

Authors (Year):	Title:	Constructs:
Kim and Biocca [209] (1997)	//	Arrival, departure
Witmer and Singer [79] (1998)—refactored in 2005 [210]	Presence Questionnaire (PQ)	Control, sensory, distraction, realism
Schubert et al. [211] (1999)	Igroup Presence Questionnaire (IPQ)	Spatial presence, involvement, realism, quality of immersion, drama, interface awareness, exploration of VE, predictability
Baños et al. [212] (2000)	Reality Judgment and Presence Questionnaire (RJ PQ)	Reality judgment, presence, emotional involvement, interaction, control, attention/flow, realism, congruence/continuity, expectations
Lessiter et al. [213] (2000)	ITC Sense Of Presence Inventory (ITC-SOPI)	Physical space, engagement, naturalness, negative effects
Lombard et al. [214] (2000)	//	Immersion, parasocial interaction, parasocial relationships, physiological response, social reality, interpersonal social richness, general social richness
Slater, Usoh, and Steed [215,216] (1998)	Slater-Usoh-Steed Questionnaire (SUS)	Sense of “being there”, locality, realism
Sas and O’Hare [217] (2003)	//	“Being there”, “not being here”, reflective consciousness
Vorderer et al. [218] (2004)	MEC Spatial Presence Questionnaire (MEC-SPQ)	Attention allocation, spatial situation model, self-location, possible actions, cognitive involvement, suspension of disbelief, domain-specific interest, visual/spatial imagery, absorption
Poels et al. [191,192] (2007)	Game Experience Questionnaire (GEQ)	Competence, sensory and imaginative immersion, flow, tension/annoyance, challenge, negative affect, positive affect
Jennett et al. [219] (2008)	Immersive Experience Questionnaire (IEQ)	Basic attention, enjoyment, transportation, temporal dissociation, challenge, emotional involvement
Cheng et al. [220] (2015)	Game Immersion Questionnaire (GIQ)	Attraction, time investment, usability, emotional attachment, decreased perceptions, presence, empathy
Makransky et al. [221] (2017)	Multimodal Presence Scale (MPS)	Physical realism, not paying attention to real environment, sense of being in the VE, unawareness of physical mediation, sense of coexistence, human realism, unawareness of social interaction artificiality, unawareness of social mediation, sense of bodily connectivity, sense of bodily extension
Khenak et al. [222] (2019)	Spatial Presence for Immersive Environments (SP-IE) Questionnaire	Spatial presence, affordance, enjoyment, reality, attention allocation, social embodiment, negative effects

authors use EEG data to confirm previous findings [239,240] regarding the activity in the parietal lobe and how it relates to the experience of presence.

6.1.4 Measuring cybersickness and VR-related discomfort

Subjective measures: Kellogg et al. [241] developed the Pensacola Motion Sickness Questionnaire (MSQ). Kennedy et al. [93] later developed a condensed version of the MSQ entitled Simulator Sickness Questionnaire (SSQ), which is the most commonly used questionnaire for evaluating cybersickness. However, as VR technology slowly begins to enter the mainstream, researchers are growing more aware of the need to differentiate between motion sickness, simulator sickness, and cybersickness, as discussed in Sect. 3. Although a popular choice among VR researchers, the SSQ may not be an ideal choice for assessing VR-induced discomfort [127,242]. Therefore, several similar questionnaires have emerged, developed specifically for the VR platform. Ames et al. [101] concluded that the SSQ did not include enough ocular symptoms to be fully appropriate for evaluating immersive environments, and developed a new questionnaire called Virtual Reality Sickness Questionnaire (VRSQ), specifically intended for use with head mounted displays. Another VR-specific SSQ-based questionnaire of the same name and abbreviation was developed thirteen years later, by Kim et al. [243]. In terms of size and questionnaire items, the VRSQ (2018) questionnaire is similar to the CyberSickness Questionnaire (CSQ) [242], also a modification of the SSQ developed for VR. A comparison of symptoms tracked by MSQ, SSQ, CSQ and both VRSQ questionnaires is presented in Table 3.

Questionnaires such as the SSQ and its variations are given after (and sometimes before) a specific VR experience. However, researchers may choose to examine the users' overall susceptibility to cybersickness, which is usually investigated prior to VR exposure, often using the revised versions [244,245] of the Motion Sickness Susceptibility Questionnaire (MSSQ) [96,246]. While the MSSQ (revised) questionnaire (long and short) investigates users' previous experiences with sickness during exposure to different types of motion, its final version does not include items pertaining to experiences that are primarily associated with visually induced sickness (e.g., virtual reality). However, considering the similarities between motion sickness and cybersickness [92], it is commonly used in VR research. Moreover, it shows a positive correlation with post-VR SSQ scores [247].

The main disadvantage of longer questionnaires is the long time it takes to complete them, especially if the study requires completing them multiple times in a session. Because of this, the time needed to fill out a multi-item questionnaire might lead to a decrease in cybersickness symptoms [248] which can influence the results. Therefore, in addition to

multiple-item questionnaires, single-item questions are also commonly used across different studies, although they tend to be more study-specific and less extensive.

In terms of measuring pain, discomfort, physical exertion and fatigue, researchers often use scales developed by Borg [249]. Due to its specific scaling, which ties verbal anchors to values between 6 and 20, the Borg Rating of Perceived Exertion (RPE) scale is able to provide an estimate of the users heart rate. The Borg CR10 scale provides a simpler scaling system, with verbal anchors corresponding to values between 0 and 10. A methodology that examines multiple symptom groups within the same study, combining the SSQ with subjective measures of ergonomic symptoms (including the use of the modified Borg CR10 scale) and digital eye strain, is presented in [127]. To gain a deeper insight regarding the effort necessary to interact with the virtual environment, and better understand which dimensions contribute towards greater frustration and fatigue, researchers may employ subjective measures of workload, such as the NASA Task Load Index (NASA-TLX) [250] or the novel Simulation Task Load Index (SIM-TLX) [251], developed with the VR platform in mind.

Objective measures: Kim et al. [252] and Dennison et al. [253] tracked various physiological signals using different modules of a Biopac polygraph: ECG, electrooculogram (EOG), electrogastogram (EGG), GSR etc. Kim et al. [252] found that gastric tachyarrhythmia, blinking, breathing and heart rate significantly correlated to the cybersickness score. Dennison et al. [253] conducted a study on twenty individuals, examining the impact of virtual reality use on cybersickness. Their findings show that changes in breathing, blinking and stomach activity may serve as indicators of cybersickness. Results by Wu et al. [254] show that impaired response inhibition can indicate cybersickness, which can be assessed by measuring inhibition-related components of event-related potentials (ERPs).

For pre-screening participants and evaluating vision-related symptoms (e.g., [138]), researchers may choose to use eye charts and tools for assessing color perception (e.g., Ishihara test [255]), distance vision (e.g., Snellen chart [256]), near vision (e.g., Fonda-Anderson chart [257]), and stereo vision acuity (e.g., Butterfly Stereo Acuity test), as well as vergence and accommodation (e.g., Royal Air Force near point rule [258]). Iskander et al. [129] mention the potential of VR HMDs equipped with eye-tracking technology, as they highlight the deficit of datasets containing captures of coordinated eye and body movement during immersive VR, which would aid in assessing visual fatigue. Eye-tracking technology enables researchers to collect different types of ocular measures, such as gaze direction, fixation duration, blink duration and frequency, and pupil dilatation. In addition to their use in assessing fatigue [36], eye movements may also be used as a measure when exploring the effect of

Table 3 Comparison of symptoms assessed by the MSQ [241], SSQ [93], VRSQ [101], CSQ [242], and VRSQ [243] questionnaires

Symptom	MSQ	SSQ	VRSQ (2005)	CSQ	VRSQ (2018)
General discomfort	X	X	X		X
Eyestrain	X	X	X	X	X
Tired eyes			X		
Sore/aching eyes			X		
Irritated eyes			X		
Watery eyes			X		
Dry eyes			X		
Hot/burning eyes			X		
Blurred vision	X	X	X	X	X
Double vision			X		
Vision discomfort			X		
Visual flashbacks	X				
Dizziness			X		
Dizziness (eyes open)	X	X		X	
Dizziness (eyes closed)	X	X		X	X
Vertigo	X	X		X	X
Headache	X	X	X	X	X
Fullness of head	X	X		X	X
Awareness of breathing	X				
Stomach awareness	X	X	X		
Decreased appetite	X				
Increased appetite	X				
Nausea	X	X	X	X	
Burping	X	X			
Vomiting	X				
Desire to move bowels	X				
Increased salivation	X	X			
Decreased salivation	X				
Sweating	X	X	X		
Fatigue	X	X	X		X
Faintness	X				
Drowsiness	X		X		
Confusion	X				
Difficulty focusing	X	X	X	X	X
Difficulty concentrating	X	X	X		
Boredom	X		X		
Depression	X				
Claustrophobia			X		
Exhilaration			X		

VR on cognitive processing, e.g., by tracking saccadic eye movements using VR-specific tools such as [259]. Moving away from ocular measures, an example of a methodology incorporating various cognitive performance (reaction time, mental rotation, visual search, visual working memory) measures is presented in [135]. Considering reaction time tests as a popular choice among tools for assessing cognitive performance, researchers may choose to use tools measuring both simple and choice reaction time, such as the Deary Liewald reaction time task [260], or a tool such as the CANTAB 5-choice reaction time task [261], which provides results for decision and motor movement speed.

7 Temporal aspects of QoE assessment

Figure 4 depicts time spans of user experience, based on models presented in [262] and [263]. Before the user even starts interacting with the system, they form a set of expectations about the experience (an *internal reference* [264]). E.g., these expectations may form as a consequence of the user's previous experience with a similar system, or they may be a result of the *halo effect*. As the user begins to interact with the system, they perform a series of momentary evaluations of the experience (comparing actual experience to their *internal reference*), based on which they are able to form a reflective evaluation of an episode of use. Repeated use of the system allows the user to make judgements over the span of multiple episodes, and impacts their summative evaluation of the system as a whole. An in-depth analysis of temporal development of QoE is given in [264].

7.1 When (and how) to measure momentary and reflective QoE

Subjective ratings of reflective QoE are usually collected post-episode via single- or multiple-item questionnaires. During use, the perceived QoE is continuously changing based on the current (momentary) level of quality, and may even increase or decrease drastically in case of sudden changes. However, as explained in [264], when an episode of use ends, and the user is inquired about their experience, they are more likely to report a level of quality that correlates to their initial/first (i.e., *the primacy effect*) or their more recent (i.e., *the recency effect*) momentary judgements, which suggests that measuring reflective QoE does not provide an accurate evaluation of momentary experience. Furthermore, if, after encountering an impairment during use, users experience a certain period in which their experience is not impaired, they may be more likely to disregard the impairment as they reflect on the experience, which is known as *the forgiveness effect* [265]. With this in mind, depending on the IFs examined in the user study, researchers should

decide whether it is more appropriate to measure momentary or reflective QoE (or both), and choose suitable measures and methods based on this decision. Researchers may ask the participants to evaluate subjective momentary QoE by assessing the quality of a series of very short (i.e., several seconds) samples which comprise a longer test stimulus, or by continuously reporting the quality of a longer stimulus using a slider or some other type of mechanism that allows for continuous collection of momentary ratings [264]. However, attempting to evaluate QoE in this way means that the user's attention is continuously being divided between the material they are trying to evaluate and the evaluation task itself [9]. In the context of user experience with a medium such as VR, which strongly relies on the sense of "being" in the virtual world, divided attention and/or constant interruptions are likely to diminish the level of presence/immersion experienced by the user [225] and thus significantly affect the overall VR QoE. A less obtrusive approach relies on the use of physiological measures with a high sampling rate [9], such as EEG, GSR, and heart rate.

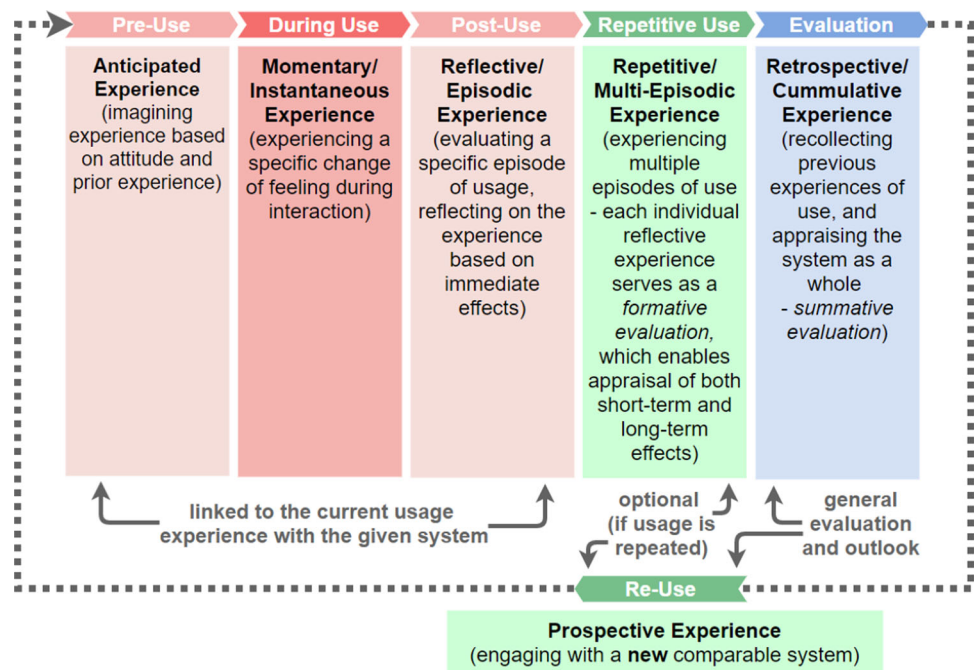
7.2 Considerations regarding the duration of a single test scenario/study session

An important issue with measuring reflective QoE is determining the optimal duration of a test scenario. ITU-T Recomm. P.809 [70], which focuses on subjective evaluation of gaming QoE, describes two testing approaches depending on the aim of the user study. A short interactive test, lasting between 90 and 120 seconds, should be adequate for assessing more straightforward QoE features (i.e., quality of interaction). Long interactive tests, usually lasting between 10 and 15 mins, are more suitable for measuring affective states and evaluating complex features such as immersion, presence, or flow. However, while aforementioned recommendations should be taken into consideration as interactive VR applications (especially VR games) share many similarities with other VEs, such as games played on less immersive platforms, researchers should also consider VR-specific issues and health risks when determining the duration of VR exposure for user studies.

Murata and Miyoshi [266] used a posturography technique based on a force platform to measure body sway during VR use. Results obtained under the control condition (i.e., while not using a computer/VR system) showed that postural instability and cybersickness tend to remain stable over the course of three hours. On the contrary, during the three-hour experiment, postural stability of participants immersed in a VR environment gradually decreased, while symptoms of cybersickness increased compared to the pre-immersion condition.

Wang and Suh [125] present a time-varying cybersickness model with trigger factors and adaptation factors, depicted in Fig. 5, and based on [267]. As users begin to experience

Fig. 4 Time spans of user experience (adapted from [262,263])



cybersickness, their body starts to adjust (see Sect. 3.5.3), leading to a decrease in cybersickness. Even though users continue to adapt to cybersickness triggers, the sensation of cybersickness has a tendency to accumulate with prolonged exposure which can eventually lead to an unenjoyable experience, although this process is slowed down when adaptation mechanisms (e.g., adjusting their movements, taking a break) are employed.

Stanney et al. [15] conducted a cybersickness study ($n = 1102$) in which participants were exposed to a virtual environment for an assigned duration (15, 30, 45 and 60 mins). The authors reported a cybersickness rate above 80% and an increase in symptom severity with longer exposure. However, around the 45 min mark nausea- and disorientation-related symptoms stopped increasing, while oculomotor symptoms continued to worsen. Longer exposures produced a greater dropout rate. During the first hour after exposure, total severity of symptoms decreased by 30.7%, but even 2 to 4 h after exposure 73% of participants were still experiencing symptoms, while 35% continued experiencing them more than 4 h after exposure. 18% of participants reported cybersickness symptoms the following morning (approx. 24 h after exposure). While reported symptoms included nausea and oculomotor symptoms, the main type of symptom that remained after a longer duration was disorientation.

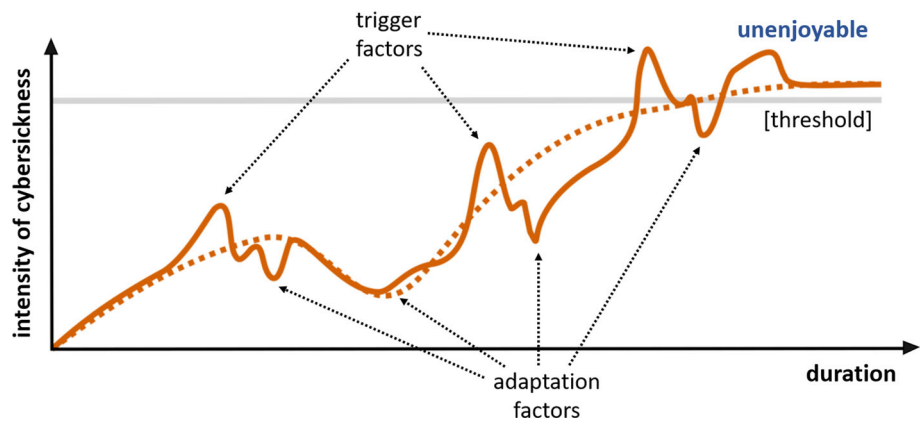
In case of multiple test scenarios during a single user session (especially if using long interactive tests), the total duration of VR exposure may significantly exceed the duration of 15 min recommended by Stanney et al. [15], or even 30 min recommended by Drachen et al. [148]. During this time, symptoms of cybersickness may accumulate.

This is an example of the effect known as *multiple treatment interference*, which happens when test subjects are asked to participate in a series of treatment conditions. In such circumstances, an effect caused by a previously experienced condition (e.g., tiredness, expertise) may carry over to the subsequent treatments, potentially influencing the results of the study [139]. In the context of VR QoE studies, *multiple treatment interference* may happen with factors/features such as physical symptoms (e.g., eye-strain, nausea), ease of use, affective states, as well as task performance measures. To a degree, randomizing the order of test scenarios may mitigate the issue of invalid QoE scores, while using test tasks that are designed with user comfort in mind (if appropriate for the study) may prevent or reduce physical symptoms. Readers interested in temporal factors involved in the experience of cybersickness may refer to [144] for a detailed overview of the topic.

7.3 Measuring repetitive and retrospective QoE

Karapanos et al. [268] discuss different approaches to collecting samples of user data in the context of repeated use. The *pre-post* approach refers to collecting and comparing participant data twice (i.e., at a point in time which is close to the beginning of the study, and again after a certain time period). The *longitudinal* approach is based on collecting a greater number of measurements. Wilson and McGill [269] highlight the deficit of longitudinal user studies evaluating the use of VR and its consequences. Considering that commercial VR is still in its early stages, there is a lack of knowledge regarding long-term usage and the way it reflects on one's

Fig. 5 Time-varying cybersickness model with trigger factors and adaptation factors (taken from [125]; based on [267])



psychological and physiological health. Aside from health related issues, examining VR use over a longer period of time is vitally important for gaining a deeper level of insight about user behaviour and preferences, and the way they change over time.

Previous research has shown that the perceived importance of different characteristics of a product (e.g., perceived stimulation [270]) tends to change over time as the novelty wears off. Additionally, Fenko et al. [271] examined the shift in sensory dominance that happens as users spend more time with the product—while vision tends to be the dominant modality in the beginning, the perceived importance of other modalities, such as touch and audition, increases with further usage. Valuable information regarding this issue in the context of VR has been provided by Bailenson and Yee [272], who conducted fifteen sessions over the course of ten weeks, observing task performance, presence, cybersickness, user behaviour, and entiativity in a collaborative virtual environment. As the study progressed, participants spent less time looking at each other, suggesting an increase in reliance on audio communication. This gradual change in behaviour, coupled with the fact that participants experienced a reduction in cybersickness with repeated use, confirms that results acquired in a single VR session are not necessarily reflective of multi-episodic VR use.

Instead of the longitudinal approach, which usually includes collecting data on multiple occasions, during or shortly after every VR use, researchers may ask participants to recall their previous VR experiences and provide their overall assessment of the system as a summative evaluation. *Retrospective* recall of a single experience, or a collection of experiences, is memory-based, and therefore may diverge from any impressions formed during or immediately after usage. However, while introducing bias, *retrospective* recall should not be dismissed in the context of evaluating user experience/acceptance, since memories condition future behavior (forming an internal reference) and, if communicated, may influence other users [268].

8 Physical environment in VR research

Even though the goal of every VR experience is to immerse the user into the virtual world (producing what we call the place illusion), the physical environment of the study remains a relevant aspect of study design. In most interactive VR applications, the user's physical movement translates to movement in a virtual environment (i.e., by moving within the tracked space, the user controls the movement of their avatar). This proves to be a safety issue as VR headsets obscure the user's view of the real world, which can potentially lead to injury and material damage. While the process of path integration (i.e. using proprioceptive cues to monitor spatial positioning) enables spatial updating in absence of visual cues [273], the focus on traversing through the virtual environment, which usually involves some degree of physical motion (turning the head and/or body, walking etc.), tends to cause disorientation with respect to the user's position in the real world. In addition to the issue of disorientation, being immersed in a virtual environment can interfere with the perception of egocentric distance [274], leading to mis-targeted movement that may result in dangerous collisions. Fortunately, the issue of incorrect egocentric distance perception has been greatly reduced in newer VR systems, such as HTC Vive [275] and Oculus Rift [276]. Nevertheless, in order to counteract these threats to participant safety, participants should be supervised at all times, and studies should be conducted in a spacious, uncluttered environment. Certain environmental conditions, such as hot temperatures or high humidity, may increase the likelihood of cybersickness [277]. Thus, it is advisable to keep the space well-ventilated, provide water and snacks [116], and a comfortable place for participants to sit or lie down in case they experience the onset of cybersickness symptoms.

Stepping away from the issue of participant comfort and safety, the location and the overall context of the experiment pose a significant influence on decisions regarding methodology, as well as on the overall outcome of the study and its internal/external validity. Due to the inherent characteristics

of the environment, and accompanying contextual variables, the study performed in a laboratory (i.e., an environment that is specifically intended for scientific research) may greatly differ from a field study (defined as “*research conducted in a place that the participant or subject perceives as a natural environment*” [139]).

8.1 In a laboratory setting

Conducting a user study in a laboratory is a very common practice in VR research, which is no surprise, given its number of benefits. Designated laboratories adapted for VR testing are usually spacious, and supplied with advanced VR equipment, which can often be problematic in terms of transportation and setup, especially if it includes large, complex devices such as a VR treadmill or exercise equipment. Conducting the study in a specialized enclosed space gives researchers more control over factors such as temperature, humidity, and the allowed number of people, which creates a higher level of comfort (both physical and psychological) compared to a public setting, while the presence of an administrator serves as an additional safety measure compared to non-supervised studies, such as those conducted in participants’ living spaces. The most obvious benefit of a laboratory environment, however, is the increased internal validity of the study, which is a result of controlled environmental variables. However, this characteristic of the laboratory setting has a downside—evaluating the application in such a sterile, artificial environment negatively impacts the external validity of the study, as acquired results may not be representative of real-world usage [139].

8.2 In “the wild”

Choosing to conduct the study outside of a laboratory requires changes in methodology and duration. These changes can go both ways - compared to laboratory studies, methodology may be more limited in case of public walk-in studies, or more extensive in case subjects are able to participate from the comfort of their homes. Likewise, study duration of field studies varies greatly—for example, a study conducted at a public place/event may have to be shortened to only a few minutes (e.g., [278]), while moving the study to a home setting may even allow for longitudinal research (e.g., [279]).

8.2.1 At a public place/event

Conducting a VR user study at a public place (e.g., amusement park, shopping mall) and/or during a public event (e.g., exhibition, convention) is a convenient way to assess short-term QoE/UX for a large number of participants. Careful selection of the venue/event can be used to facilitate access to the target demographic (e.g., researchers may choose to

conduct a gaming QoE/UX study at a gaming convention visited by a large number of avid gamers) without a tedious pre-screening process.

However, this type of setting has its fair share of obstacles in terms of methodology design. Firstly, in such cases, the duration of the study process is generally kept at a bare minimum (e.g., 2–5 min [278]). Due to such brief exposure to VR, participants are not likely to experience more complex aspects of VR experience (e.g., immersion [70], cybersickness) to their fullest degree. Thus, when conducting a study in such a public scenario, researchers should be aware of the limitations imposed on the choice of observed factors/features, and their implications on the validity of subsequent results. Wearable devices, whether used for position tracking or collecting physiological signals, are generally too cumbersome for the fast subject turnover of a public walk-in study. For example, there may not be enough time to calibrate devices for individual use, or acquire baseline measurements of physiological signals—e.g., each analysis interval during which continuous EEG data is being collected should be around 5–10 min long, following a 2–5 min period for the collection of baseline measurements [280]. Therefore, researchers may choose to use questionnaires [281], or rely only on behavioral methods [278]. Moreover, using a VR application in a public setting could trigger certain users to feel uncomfortable, exposed, or pressured, as certain people consider the public use of VR to be embarrassing [152,282,283], which may influence their subjective assessment, or cause them to interact with the application differently than they would if they were using it in a more private situation. The influence of being watched whilst immersed in VR is analyzed in detail in a paper by Mai et al. [283]. The authors also elaborate on other possible issues with the public use of VR, such as unwanted touches and the increased likelihood of injury in case of collision with a bystander. Based on these observations, the authors present valuable findings and suggestions on the use of spatial, visual and auditory separation between the person immersed in a VR experience and other people, the inclusion of a supervising person to watch over the user and help them feel more comfortable, and scenario/methodology design that allows the user to slowly ease into the VR experience without feeling too self-conscious. Additional guidelines on how to provide a more comfortable experience for participants using VR in public are presented in [284].

8.2.2 At the target location

Depending on the intended goal, a VR application may be developed for personal use, or as an education/visualization tool. By moving the study to the target location (i.e., conducting a field study) such as a school, or a living space, researchers can avoid the negative impact of artificial labo-

ratory setting on the external validity of the study, therefore achieving a higher level of experimental realism [139]. Using the examined system, service or application in an environment that is perceived as more natural mitigates the issue of participant reactivity (i.e., display of modified participant behaviours resulting from the awareness that they are being observed/tested [139]). However, while improving external validity, moving away from the sterile laboratory environment tends to decrease internal validity, as it becomes harder to control environmental variables of the study [139]. Additionally, based on our experience in conducting VR user studies, we would argue that providing a large number of participants/locations with expensive VR equipment, lending devices for a longer period of time, or transporting complicated VR setups to a target location may be problematic for the institution conducting the study, which greatly impacts the scope of the field study in terms of used devices, population size, study duration, observed IFs, etc. The same availability issues hold true for the use of devices for physiological data collection. Thus, researchers may have to rely solely on self-reported data.

A large percentage of VR applications is intended for personal use in a private space, but the option of conducting VR studies from the users' homes is especially important to consider in the context of current events regarding the COVID-19 pandemic, as researchers struggle with limited access to laboratories and public spaces, and hygiene concerns regarding shared VR HMD use. While evaluating the use of VR in home conditions is slowly becoming more achievable, as the number of casual VR users has started to increase over the last several years [269], VR owners are still a definite minority. When conducting from-home studies, the majority of participants will have to be provided with the necessary equipment, which, as discussed, tends to be highly impractical and/or financially straining for research institutions conducting the experiment, especially in terms of more advanced VR systems. Depending on the goal of the study, a more achievable solution may be to focus on mobile VR, which is less expensive, standalone (i.e., does not rely on a VR-ready computer), and easy to set up. However, a more promising solution for this issue may be found in the use of crowdsourcing for QoE assessment [143], leveraging Internet platforms for the recruitment of VR owners for participation in online studies.

For instance, Steed et al. [179] conducted a field study exploring presence and embodiment in immersive VR using mobile VR platforms. With respect to limitations of popular mobile VR devices, the content was designed to be non-interactive as a way to mitigate the chance of injury and adverse reactions, as this was a public, non-supervised study. The application was distributed via app stores. While this type of distribution made it available to a large number of customers, data collection (answers to two questionnaires, device information, head-tracking information) was

performed only in case of consenting users. The benefit of this method of gathering participants is the broadening of the population set in comparison to a typical laboratory study. However, the authors note that the issue of test material design is more relevant in case it is being distributed in such a public way (compared to laboratory studies), considering that it has to rely on visual attractiveness and content quality in order to stand out from other VR applications, while avoiding elements that may provoke a stressful response due to ethical reasons. Recent examples of the use of crowdsourcing in VR user research are presented in [127,285]. However, in cases where an official supervisor is not present to monitor the use of an application, participants should still be monitored (e.g., by a family member, friend, or colleague) to prevent injury.

9 Summary of key challenges

9.1 Identifying influence factors and features to be used for assessing and modeling QoE

As listed in Sect. 3, there are multiple factors responsible for the formation of VR user experience. Some of those factors are relevant to other types of audio-visual services, even non-interactive ones, while others are specific to immersive interactive VR. By identifying key factors and examining their influence on different QoE features, researchers are able to make adjustments to their study design, and collect data to be used for QoE modeling. Careful consideration of the VR market, especially in light of recent findings pertaining to VR user acceptance, helps with narrowing the focus towards most relevant aspects of the VR experience. Recent technological advancements regarding commercial VR technology, as well as the arrival of 5G networks, highlight the need for further research observing the influence of system IFs on the overall QoE.

In addition to its immersivity, VR use is characterized by higher levels of discomfort compared to less intrusive platforms. Researchers should consider examining VR-induced discomfort as an important feature to incorporate into UX/QoE and technology acceptance models. Additionally, while the existing body of work addressing cybersickness is relatively large, recent findings call for a shift towards exploring other symptoms of discomfort (i.e., digital eye strain, ergonomic factors related to headset design, control modality, and interaction) and fatigue, as well as cognitive performance aftereffects of VR use.

9.2 Defining the test methodology

While it is unethical to expose participants to situations that may bring significant or long-term psychological or physiological harm, some researchers argue that it is necessary

to include more sensitive populations in VR research, as it provides valuable information which can be used to adapt existing systems to their specific needs. Therefore, there is a need for guidelines regarding pre-screening methods and exclusion criteria, with respect to ethical issues. Considering university students are generally over-represented in user research, efforts should be made towards broadening the participant population in future studies. However, this calls for additional research and methodology guidelines pertaining to more sensitive demographics, such as children and the elderly. Additionally, researchers should focus on including participants with various levels of experience, and work towards determining the impact of the novelty effect on the overall QoE.

Standardized test material facilitates comparison between studies, as well as the reproduction of study results. Therefore, efforts should be invested towards designing test applications which include relevant interaction methods, and are suitable for use with different types of I/O devices.

Subjective methods are commonly used in VR user studies. While most researchers use paper questionnaires, incorporating questionnaires into VEs used for testing should be encouraged. Even though they offer valuable information, subjective methods should be combined with objective methods for more relevant results, and their mutual relationship should be examined. Even though there is a significant number of commonly used questionnaires (mostly related to cybersickness and presence), there is still room for improvement with regards to addressing specific dimensions of VR use. In terms of objective measures, new technology (e.g., eye-tracking in VR HMDs) facilitates the development of novel methods for assessing user experience.

Considering VR is generally more physically exhausting and more likely to induce cybersickness compared to most other platforms, the recommended duration of each episode of use is up for debate, with certain sources recommending time frames as low as 15 min. Unfortunately, performing a user study while limiting VR exposure to such a short duration also limits the validity of its results, as they may not be representative of realistic, long-term use. Therefore, there is a need for guidelines addressing study duration. Additionally, there is currently a deficit of studies exploring the effect of prolonged VR exposure, as well as a deficit of longitudinal VR studies. On a similar note, the majority of QoE studies is conducted in a sterile laboratory environment. Along with extending the observed time frame, conducting research in a more realistic setting is likely to result in valuable insights and greater external validity. Recent events and regulations related to the COVID-19 pandemic highlight the importance of considering the use of crowdsourcing to facilitate VR user research.

10 Conclusion

In this paper we have provided an overview of perception-based QoE assessment for interactive VR applications, organized into sections discussing the motivation behind VR user research, relevant IFs and QoE features, pre-screening and participant choice, test material, subjective and objective measures, as well as study duration and preparation of appropriate study environment. Guided by the multimodality of the VR platform, along with its wide array of potential uses, we have based this overview on sources stemming from various branches of science. Bringing together key findings from literature and existing standards, we have presented a collection of resources, explanations, and recommendations to serve as a reference for academic and industry researchers interested in conducting VR user studies. Based on our findings, we have summarized key challenges to be addressed in future research: identifying IFs and features to be used for QoE modeling, as well as addressing different ethical and practical aspects of methodology design for VR user research. We note, however, that each of the presented elements of perception-based QoE assessment requires its own in-depth review, as the aim of this paper was to provide only a concise, high-level overview of the topic.

Funding This work has been fully supported by the Croatian Science Foundation under the project Modeling and Monitoring QoE for Immersive 5G-Enabled Multimedia Services (Q-MERSIVE), grant numbers IP-2019-04-9793 and DOK-2020-01-3779.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Pallavicini F, Ferrari A, Zini A, Garcea G, Zancacchi A, Barone G, Mantovani F (2017) What Distinguishes a traditional gaming experience from one in virtual reality? An exploratory study. In: International conference on applied human factors and ergonomics, pp 225–231. Springer
2. Pallavicini F, Pepe A, Minissi ME (2019) Gaming in virtual reality: what changes in terms of usability, emotional response and sense of presence compared to non-immersive video games? *Simul Gam* 50(2):136–159
3. Shelstad WJ, Smith DC, Chaparro BS (2017) Gaming on the rift: how virtual reality affects game user satisfaction. In: Proceedings of the human factors and ergonomics society annual meeting, vol 61(1). SAGE Publication, Sage CA: Los Angeles, CA, pp 2072–2076

4. Steam (2021) Steam hardware & software survey. <https://store.steampowered.com/hwsurvey/>. Accessed 6 Apr 2021
5. Le Callet P, Möller S, Perkis A et al (2012) Qualinet white paper on definitions of quality of experience. European Network on Quality of Experience in Multimedia Systems and Services 3
6. Perkis A, Timmerer C et al (2020) QUALINET white paper on definitions of immersive media experience (IMEx). European Network on Quality of Experience in Multimedia Systems and Services, 14th QUALINET meeting (online)
7. Lavoie R, Main K, King C, King D (2021) Virtual experience, real consequences: the potential negative emotional consequences of virtual reality gameplay. *Virtual Real* 25(1):69–81
8. Caserman P, Garcia-Agundez A, Zerban AG, Göbel S (2021) Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. *Virtual Real* 25:1153–1170
9. Barreda-Ángeles M, Redondo-Tejedor R, Pereda-Baños A (2018) Psychophysiological methods for quality of experience research in virtual reality systems and applications. *IEEE COMSOC MMTC Commun Front* 4(1):14–20
10. Parés N, Parés R (2006) Towards a model for a virtual reality experience: the virtual subjectiveness. *Presence* 15(5):524–538
11. Aukstakalnis S (2016) Practical augmented reality: a guide to the technologies, applications, and human factors for AR and VR. Addison-Wesley Prof
12. Liberatore MJ, Wagner WP (2021) Virtual, mixed, and augmented reality: a systematic review for immersive systems research. *Virtual Real* 25:773–799
13. Yao S-H, Fan C-L, Hsu C-H (2019) Towards quality-of-experience models for watching 360 videos in head-mounted virtual reality. In: 2019 11th international conference on quality of multimedia experience, pp 1–3. IEEE
14. Steuer J (1992) Defining virtual reality: dimensions determining telepresence. *J Commun* 42(4):73–93
15. Stanney KM, Hale KS, Nahmens I, Kennedy RS (2003) What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. *Hum Fact* 45(3):504–520
16. Weech S, Kenny S, Barnett-Cowan M (2019) Presence and cybersickness in virtual reality are negatively related: a review. *Front Psychol* 10:158
17. Raake A, Egger S (2014) Quality and quality of experience. In: *Quality of experience*. Springer, pp 11–33
18. ITU-T Work Item P.QXM (SG 12) (2021) QoE Assessment of eXtended Reality (XR) meetings. https://www.itu.int/itu-t/workprog/wp_item.aspx?isn=15113, Accessed 5 Nov 2021
19. Wechsung I, De Moor K (2014) Quality of experience versus user experience. In: *Quality of experience*. Springer, pp 35–54
20. Robertson DW (1946) A note on the classical origin of circumstances in the medieval confessional. *Stud Philol* 43(1):6–14
21. Coie P (2020) Augmented and virtual reality survey report. <https://www.perkinscoie.com/images/content/2/3/v4/231654/2020-AR-VR-Survey-v3.pdf>, Accessed 21 Apr 2020
22. Vega MT, Liaskos C, Abadal S, Papapetrou E, Jain A, Mouhouche B, Kalem G, Ergüt S, Mach M, Sabol T et al (2020) Immersive interconnected virtual and augmented reality: a 5G and IoT perspective. *J Netw Syst Manag* 28(4):796–826
23. Sagnier C, Loup-Escande E, Lourdeaux D, Thouvenin I, Valléry G (2020) User acceptance of virtual reality: an extended technology acceptance model. *Int J Human Comput Interact* 36:993–1007
24. Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q* 13:319–340
25. Hassenzahl M (2008) User experience (UX) towards an experiential perspective on product quality. In: *Proceedings of the 20th conference on l'interaction homme-machine*, pp 11–15
26. Manis KT, Choi D (2019) The virtual reality hardware acceptance model (VR-HAM): extending and individualizing the technology acceptance model (TAM) for virtual reality hardware. *J Bus Res* 100:503–513
27. Whalen TE, Noël S, Stewart J (2003) Measuring the human side of virtual reality. In: *IEEE international symposium on virtual environments, human-computer interfaces and measurement systems*. IEEE, pp 8–12
28. ITU-T Recomm. G.1035 (2020) Influencing factors on quality of experience for virtual reality services
29. Hupont I, Gracia J, Sanagustin L, Gracia MA (2015) How do new visual immersive systems influence gaming QoE? A use case of serious gaming with Oculus Rift. In: *2015 7th international workshop on quality of multimedia experience*. IEEE, pp 1–6
30. Meline T (2009) A research primer for communication sciences and disorders. Allyn & Bacon
31. Chihara T, Seo A (2018) Evaluation of physical workload affected by mass and center of mass of head-mounted display. *Appl Ergon* 68:204–212
32. Laffont P-Y, Martin T, Gross M, De Tan W, Lim CT, Au A, Wong R (2016) Rectifeye: a vision-correcting system for virtual reality. In: *SIGGRAPH ASIA 2016 VR showcase*. Association for Computing Machinery, pp 1–2
33. Suznjevic M, Mandurov M, Matijasevic M (2017) Performance and QoE assessment of HTC Vive and Oculus Rift for pick-and-place tasks in VR. In: *2017 9th international conference on quality of multimedia experience*. IEEE, pp 1–3
34. Gonçalves G, Monteiro P, Melo M, Vasconcelos-Raposo J, Bessa M (2020) A comparative study between wired and wireless virtual reality setups. *IEEE Access* 8:29249–29258
35. Soler-Dominguez JL, Camba JD, Contero M, Alcañiz M (2017) A proposal for the selection of eye-tracking metrics for the implementation of adaptive gameplay in virtual reality based games. In: *International conference on virtual, augmented and mixed reality*. Springer, pp 369–380
36. Wang Y, Zhai G, So Chen, Min X, Gao Z, Song X (2019) Assessment of eye fatigue caused by head-mounted displays using eye-tracking. *Biomed Eng Online* 18(1):1–19
37. Patney A, Salvi M, Kim J, Kaplanyan A, Wyman C, Benty N, Luebke D, Lefohn A (2016) Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans Graph* 35(6):1–12
38. Hsu C-F, Chen A, Hsu C-H, Huang C-Y, Lei C-L, Chen K-T (2017) Is foveated rendering perceivable in virtual reality? Exploring the efficiency and consistency of quality assessment methods. In: *Proceedings of the 25th ACM international conference on multimedia*, pp 55–63
39. Ma X, Yao Z, Wang Y, Pei W, Chen H (2018) Combining brain-computer interface and eye tracking for high-speed text entry in virtual reality. In: *23rd international conference on intelligent user interfaces*, pp 263–267
40. Hameed A, Perkis A, Möller S (2021) Evaluating hand-tracking interaction for performing motor-tasks in VR learning environments. In: *2021 13th international conference on quality of multimedia experience (QoMEX)*. IEEE, pp 219–224
41. Raaen K, Kjellmo I (2015) Measuring latency in virtual reality systems. In: *International conference on entertainment computing*. Springer, pp 457–462
42. Albert R, Patney A, Luebke D, Kim J (2017) Latency requirements for foveated rendering in virtual reality. *ACM Trans Appl Percept* 14(4):1–13
43. Brunnström K, Dima E, Qureshi T, Johanson M, Andersson M, Sjöström M (2020) Latency impact on quality of experience in a virtual reality simulator for remote control of machines. *Signal Process Image Commun* 89:116005
44. Batmaz AU, Machuca MDB, Pham DM, Stuerzlinger W (2019) Do head-mounted display stereo deficiencies affect 3D pointing

- tasks in AR and VR? In: 2019 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, pp 585–592
45. Mehrfard A, Fotouhi J, Taylor G, Forster T, Navab N, Fuerst B (2019) A comparative analysis of virtual reality head-mounted display systems. arXiv preprint [arXiv:1912.02913](https://arxiv.org/abs/1912.02913)
 46. Jerald J (2015) *The VR Book: human-centered design for virtual reality*. Morgan & Claypool
 47. Doumanoglou A, Griffin D, Serrano J, Zioulis N, Phan TK, Jiménez D, Zarpalas D, Alvarez F, Rio M, Daras P (2018) Quality of experience for 3-D immersive media streaming. *IEEE Trans Broadcast* 64(2):379–391
 48. Vlahovic S, Suznjevic M, Skorin-Kapov L (2019) Challenges in assessing network latency impact on QoE and in-game performance in VR first person shooter games. In: *ConTEL 2019*. IEEE, pp 1–8
 49. Concannon D, Flynn R, Murray N (2019) A quality of experience evaluation system and research challenges for networked virtual reality-based teleoperation applications. In: *Proceedings of the 11th ACM workshop on immersive mixed and virtual environment systems*, pp 10–12
 50. Sharma SK, Woungang I, Anpalagan A, Chatzinotas S (2020) Toward tactile internet in beyond 5G Era: recent advances, current issues, and future directions. *IEEE Access* 8:56948–56991
 51. Fan C-L, Yen S-C, Huang C-Y, Hsu C-H (2020) On the optimal encoding ladder of tiled 360 videos for head-mounted virtual reality. *IEEE Trans Circuits Syst Video Technol* 31(4):1632–1647
 52. Zou W, Feng S, Mao X, Yang F, Ma Z (2021) Enhancing quality of experience for cloud virtual reality gaming: an object-aware video encoding. In: *2021 IEEE international conference on multimedia & expo workshops (ICMEW)*. IEEE, pp 1–6
 53. Mai X, Li C, Zhang S, Le CP (2020) State-of-the-art in 360 video/image processing: perception, assessment and compression. *IEEE J Sel Topics Signal Process* 14(1):5–26
 54. ITU-T Recomm. P.919 (2020) Subjective test methodologies for 360° video on head-mounted displays
 55. Gutierrez J, et al (2021) Subjective evaluation of visual quality and simulator sickness of short 360 videos: ITU-T Rec. P. 919. *IEEE Trans Multimed*
 56. Narbutt M, O’Leary S, Allen A, Skoglund J, Hines A (2017) Streaming VR for immersion: quality aspects of compressed spatial audio. In: *2017 23rd international conference on virtual system & multimedia (VSMM)*. IEEE, pp 1–6
 57. Schwarz S, Preda M, Baroncini V, Budagavi M, Cesar P, Chou PA, Cohen RA, Krivokuća M, Lasserre S, Li Z et al (2018) Emerging mpeg standards for point cloud compression. *IEEE J Emerg Sel Topics Circuits Syst* 9(1):133–148
 58. Graziosi D, Nakagami O, Kuma S, Zaghetto A, Suzuki T, Tabatabai A (2020) An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC). *APSIPA Trans Signal Inf Process* 9:E13
 59. Roth D, Lugrin J-L, Galakhov D, Hofmann A, Bente G, Latoschik ME, Fuhrmann A (2016) Avatar realism and social interaction quality in virtual reality. In: *2016 IEEE virtual reality (VR)*. IEEE, pp 277–278
 60. Garau M, Slater M, Vinayagamoorthy V, Brogni A, Steed A, Angela SM (2003) The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp 529–536
 61. Hvass J, Larsen O, Vendelbo K, Nilsson N, Nordahl R, Serafin S (2017) Visual realism and presence in a virtual reality game. In: *2017 3DTV conference: the true vision-capture, transmission and display of 3D video (3DTV-CON)*. IEEE, pp 1–4
 62. Vlahović S, Suznjevic M, Skorin-Kapov L (2018) Subjective assessment of different locomotion techniques in virtual reality environments. In: *2018 tenth international conference on quality of multimedia experience (QoMEX)*. IEEE, pp 1–3
 63. Boletsis C, Cedergren JE (2019) VR locomotion in the new era of virtual reality: an empirical comparison of prevalent techniques. *Adv Human Comput Interact* 2019:7420781
 64. Hameed A, Perkis A (2018) Spatial storytelling: finding interdisciplinary immersion. In: *International conference on interactive digital storytelling*. Springer, pp 323–332
 65. Irshad S, Perkis A (2020) Increasing user engagement in virtual reality: the role of interactive digital narratives to trigger emotional responses. In: *Proceedings of the 11th Nordic conference on human-computer interaction: shaping experiences, shaping society*, pp 1–4
 66. Paes D, Irizarry J (2018) A usability study of an immersive virtual reality platform for building design review: considerations on human factors and user interface. In: *Construction research congress*, vol 2018
 67. Wang Y, Hu Y, Chen Y (2021) An experimental investigation of menu selection for immersive virtual environments: fixed versus handheld menus. *Virtual Real* 25(2):409–419
 68. ITU-T Recomm. G.1032 (2017) Influence factors on gaming quality of experience. International Telecommunication Union-Telecommunication Standardization Sector
 69. Christensen JV, Mathiesen M, Poulsen JH, Ustrup EE, Kraus M (2018) Player experience in a VR and non-VR multiplayer game. In: *Proceedings of the virtual reality international conference-laval virtual*. ACM, p 10
 70. ITU-T Recomm. P.809 (2018) Subjective evaluation methods for gaming quality. International Telecommunication Union-Telecommunication Standardization Sector
 71. Möller S, Schmidt S, Beyer J (2013) Gaming taxonomy: an overview of concepts and evaluation methods for computer gaming QoE. In: *2013 5th international workshop on quality of multimedia experience (QoMEX)*. IEEE, pp 236–241
 72. Jekosch U (2006) *Voice and speech quality perception: assessment and evaluation*. Springer, Berlin
 73. Möller S, Wältermann M, Garcia M-N (2014) Features of quality of experience. In: *Quality of experience*. Springer, pp 73–84
 74. Urvoy M, Barkowsky M, Le Callet P (2013) How visual fatigue and discomfort impact 3D-TV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors. *Ann Telecommun* 68(11):641–655
 75. Schuemie MJ, Van Der Straaten P, Krijn M, Van Der Mast CAPG (2001) Research on presence in virtual reality: a survey. *Cyber Psychol Behav* 4(2):183–201
 76. Slater M, Usoh M (2013) An experimental exploration of presence in virtual environments. Technical report, Queen Mary University of London, London, UK
 77. Lee KM (2004) Presence. *Explic Commun Theory* 14(1):27–50
 78. Durlach N, Slater M (2000) Presence in shared virtual environments and virtual togetherness. *Presence Teleoper Virtual Environ* 9(2):214–217
 79. Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. *Presence* 7(3):225–240
 80. Slater M, Wilbur S (1997) A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence Teleoper Virtual Environ* 6(6):603–616
 81. Cummings JJ, Bailenson JN (2016) How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychol* 19(2):272–309
 82. Wirth W, Hartmann T, Böcking S, Vorderer P, Klimmt C, Schramm H, Saari T, Laarni J, Ravaja N, Gouveia FR et al (2007) A process model of the formation of spatial presence experiences. *Media Psychol* 9(3):493–525

83. Balakrishnan B, Shyam Sundar S (2011) Where am I? How can I get there? Impact of navigability and narrative transportation on spatial presence. *Human Comput Interact* 26(3):161–204
84. Skarbez R, Brooks FP Jr, Whitton MC (2018) A survey of presence and related concepts. *ACM Comput Surv* 50(6):96
85. Weibel D, Wissmath B, Mast FW (2010) Immersion in mediated environments: the role of personality traits. *Cyberpsychol Behav Soc Netw* 13(3):251–256
86. Slater M, Sanchez-Vives MV (2014) Transcending the self in immersive virtual reality. *Computer* 47(7):24–30
87. Waltemate T, Gall D, Roth D, Botsch M, Latoschik ME (2018) The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans Vis Comput Graph* 24(4):1643–1652
88. Yee N, Bailenson J (2007) The proteus effect: the effect of transformed self-representation on behavior. *Human Commun Res* 33(3):271–290
89. Kennedy RS, Berbaum KS, Drexler J (1994) Methodological and measurement issues for identification of engineering features contributing to virtual reality sickness. In: *Proceedings of image VII conference*, Tucson, AZ
90. Money KE (1970) Motion sickness. *Phys Rev* 50(1):1–39
91. Kennedy RS, Hettinger LJ, Lilienthal MG (1988) Simulator sickness. *Motion Space Sick*, pp 317–341
92. Gavvani AM, Walker FR, Hodgson DM, Nalivaiko E (2018) A comparative study of cybersickness during exposure to virtual reality and classic motion sickness: are they different? *J Appl Phys* 125(6):1670–1680
93. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int J Aviat Psychol* 3(3):203–220
94. Stanney KM, Kennedy RS, Drexler JM (1997) Cybersickness is not simulator sickness. In: *Proceedings of the human factors and ergonomics society annual meeting* 41(2):1138–1142
95. LaViola JJ Jr (2000) A discussion of cybersickness in virtual environments. *ACM Sigchi Bull* 32(1):47–56
96. Reason JT, Brand JJ (1975) *Motion sickness*. Academic Press
97. Riccio GE, Stoffregen TA (1991) An ecological theory of motion sickness and postural instability. *Ecol Psychol* 3(3):195–240
98. Treisman M (1977) Motion sickness: an evolutionary hypothesis. *Science* 197(4302):493–495
99. Geršak G, Huimin L, Guna J (2020) Effect of VR technology matureness on VR sickness. *Multimed Tools Appl* 79(21):14491–14507
100. Papadakis G, Mania K, Koutroulis E (2011) A system to measure, control and minimize end-to-end head tracking latency in immersive simulations. In: *Proceedings of the 10th international conference on virtual reality continuum and its applications in industry*. ACM, pp 581–584
101. Ames SL, Wolffsohn JS, McBrien NA (2005) The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. *Optom Vis Sci* 82(3):168–176
102. Adam JJ, Krum DM, Bolas M (2014) The effect of eye position on the view of virtual geometry. In: *2014 IEEE virtual reality*. IEEE, pp 87–88
103. Harwood K, Foley P (1987) Temporal resolution: an insight into the video display terminal (VDT) problem. *Hum Factor* 29(4):447–452
104. Landis C (1954) Determinants of the critical flicker-fusion threshold. *Physiol Rev* 34(2):259–286
105. Frank Biocca (1992) Will simulation sickness slow down the diffusion of virtual environment technology? *Presence Teleoper Virtual Environ* 1(3):334–343
106. Primeau G (2000) Wide-field-of-view SVGA sequential color HMD for use in anthropomorphic telepresence applications. In: *Helmet-and head-mounted displays V*. vol 4021. International Society for Optics and Photonics, pp 11–19
107. Kolasinski EM (1995) Simulator sickness in virtual environments. Technical report, Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA
108. Lin JJ-W, Duh HB-L, Parker DE, Abi-Rached H, Furness TA (2002) Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In: *Proceedings of IEEE virtual reality 2002*, pp 164–171. IEEE
109. Bonato F, Bubka A, Palmisano S, Phillip D, Moreno G (2008) Vection change exacerbates simulator sickness in virtual environments. *Presence Teleoper Virtual Environ* 17(3):283–292
110. Stanney KM, Hash P (1998) Locus of user-initiated control in virtual environments: influences on cybersickness. *Presence* 7(5):447–459
111. Porcino TM, Clua E, Trevisan D, Vasconcelos CN, Valente L (2017) Minimizing cyber sickness in head mounted display systems: design guidelines and applications. In: *2017 IEEE 5th international conference on serious games and applications for health*, pp 1–6. IEEE
112. Silva BM, Fernando P (2019) Early prediction of cybersickness in virtual, augmented & mixed reality applications: a review. In: *2019 IEEE 5th international conference for convergence in technology*, pp 1–6
113. Hildebrandt J, Schmitz P, Valdez AC, Kobbelt L, Ziefle M (2018) Get well soon! Human factors' influence on cybersickness after redirected walking exposure in virtual reality. In: *International conference on virtual, augmented and mixed reality*. Springer, pp 82–101
114. Kim H, Kim DJ, Chung WH, Park K-A, Kim JDK, Kim D, Kim K, Jeon HJ (2021) Clinical predictors of cybersickness in virtual reality (VR) among highly stressed people. *Sci Rep* 11(1):1–11
115. Stanney KM, Kingdon KS, Graeber D, Kennedy RS (2002) Human performance in immersive virtual environments: effects of exposure duration, user control, and scene complexity. *Hum Perform* 15(4):339–366
116. Brooks JO, Goodenough RR, Crisler MC, Klein ND, Alley RL, Koon BL, Logan WC Jr, Ogle JH, Tyrrell RA, Wills RF (2010) Simulator sickness during driving simulation studies. *Accid Anal Prev* 42(3):788–796
117. Munafo J, Diedrick M, Stoffregen TA (2017) The virtual reality head-mounted display oculus rift induces motion sickness and is sexist in its effects. *Exp Brain Res* 235(3):889–901
118. Barratt MR, Pool SL (2008) *Principles of clinical medicine for space flight*. Springer, Berlin
119. Johnson DM (2007) Simulator sickness research summary. Technical report. Army Research Institute for the Behavioral and Social Sciences, Fort Rucker, AL
120. Nichols S, Ramsey AD, Cobb S, Neale H, D'Cruz M, Wilson JR (2000) Incidence of virtual reality induced symptoms and effects (VRISE) in desktop and projection screen display systems. HSE Contract Research Report
121. Kennedy RS, Berbaum KS, Lilienthal MG, Dunlap WP, Mulligan BE (1987) Guidelines for alleviation of simulator sickness symptomatology. Technical report, Naval Training Systems Center, Orlando, FL
122. Frank Lawrence H, Kennedy Robert S, McCauley ME, Root RW, Kellogg RS (1984) Simulator sickness: sensorimotor disturbances induced in flight simulators. Technical report, Naval Training Systems Center, Orlando, FL
123. Young SD, Adelstein BD, Ellis SR (2007) Demand characteristics in assessing motion sickness in a virtual environment: or does taking a motion sickness questionnaire make you sick? *IEEE Trans Vis Comput Graph* 13(3):422–428
124. Farmer AD, Ban VF, Coen SJ, Sanger GJ, Barker GJ, Gresty MA, Giampietro VP, Williams SC, Webb DL, Hellström PM et al

- (2015) Visually induced nausea causes characteristic changes in cerebral, autonomic and endocrine function in humans. *J Physiol* 593(5):1183–1196
125. Wang G, Suh A (2019) User adaptation to cybersickness in virtual reality: a qualitative study. In: 27th European conference on information systems
 126. Vlahovic S, Suznjec M, Pavlin-Bernardic N, Skorin-Kapov L (2021) The effect of VR gaming on discomfort, cybersickness, and reaction time. In: 2021 13th international conference on quality of multimedia experience (QoMEX). IEEE, pp 163–168
 127. Hirzle T, Cordts M, Rukzio E, Gugenheimer J, Bulling A (2021) A critical assessment of the use of ssq as a measure of general discomfort in vr head-mounted displays. In: Proceedings of the 2021 ACM CHI virtual conference on human factors in computing systems-CHI, Yokohama, Japan, pp 8–13
 128. Kooi FL, Toet A (2004) Visual comfort of binocular and 3D displays. *Displays* 25(2–3):99–108
 129. Iskander J, Hossny M, Nahavandi S (2018) A review on ocular biomechanic models for assessing visual fatigue in virtual reality. *IEEE Access* 6:19345–19361
 130. Shibata T (2002) Head mounted display. *Displays* 23(1–2):57–64
 131. Hua H (2017) Enabling focus cues in head-mounted displays. *Proc IEEE* 105(5):805–824
 132. Rash CE, McLean WE, Mozo BT, Licina JR, McEntire BJ (1999) Human factors and performance concerns for the design of helmet-mounted displays. In: RTO HFM symposium on current aeromedical issues in rotary wing operation
 133. Knight JF, Baber C (2007) Effect of head-mounted displays on posture. *Hum Factors* 49(5):797–807
 134. Jang S, Stuerzlinger W, Ambike S, Ramani K (2017) Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion. In: Proceedings of the 2017 CHI conference on human factors in computing systems, pp 3328–3339
 135. Mittelstaedt JM, Wacker J, Stelling D (2019) VR aftereffect and the relation of cybersickness and cognitive performance. *Virtual Real* 23(2):143–154
 136. Nalivaiko E, Davis SL, Blackmore KL, Vakulin A, Nesbitt KV (2015) Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiol Behav* 151:583–590
 137. Nesbitt K, Davis S, Blackmore K, Nalivaiko E (2017) Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays* 48:1–8
 138. Szpak A, Michalski SC, Loetscher T (2020) Exergaming with beat saber: an investigation of virtual reality aftereffects. *J Med Internet Res* 22(10):e19840
 139. Gravetter FJ, Forzano L-AB (2018) Research methods for the behavioral sciences. Cengage Lear
 140. Thorndike EL (1920) A constant error in psychological ratings. *J Appl Psychol* 4(1):25–29
 141. Minge M, Thüring M (2018) Hedonic and pragmatic halo effects at early stages of user experience. *Int J Hum Comput Stud* 109:13–25
 142. Fairchild KM, Lee BH, Loo J, Ng H, Serra L (1993) The heaven and earth virtual reality: designing applications for novice users. In: Proceedings of IEEE virtual reality annual international symposium. IEEE, pp 47–53
 143. Hossfeld T, Keimel C, Timmerer C (2014) Crowdsourcing quality-of-experience assessments. *Computer* 47(9):98–102
 144. Duzmańska N, Strojny P, Strojny A (2018) Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Front Psychol* 9:2132
 145. Madary M, Metzinger TK (2016) Real virtuality: a code of ethical conduct. Recommendations for good scientific practice and the consumers of VR-Technology. *Front Robot AI* 3:3
 146. Behr K-M, Nosper A, Klimmt C, Hartmann T (2005) Some practical considerations of ethical issues in VR research. *Presence* 14(6):668–676
 147. Lewis CH, Griffin MJ (1997) Human factors consideration in clinical applications of virtual reality. *Stud Health Technol Inform* 44:35–58
 148. Drachen A, Mirza-Babaei P, Nacke LE (2018) Games user research. Oxford University Press
 149. Felnhofer A, Kothgassner OD, Beutl L, Hlavacs H, Kryspin-Exner I (2012) Is virtual reality made for men only? Exploring gender differences in the sense of presence. In: Proceedings of the international society on presence research, pp 103–112
 150. Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? *Behav Brain Sci* 33(2–3):61–83
 151. Plechatá A, Sahula V, Fayette D, Fajnerová I (2019) Age-related differences with immersive and non-immersive virtual reality in memory assessment. *Front Psychol* 10:1330
 152. Liu Q, Wang Y, Tang Q, Liu Z (2020) Do you feel the same as I do? Differences in virtual reality technology experience and acceptance between elderly adults and college students. *Front Psychol* 11:2555
 153. Serino S, Scarpina F, Dakanalis A, Keizer A, Pedroli E, Castellnuovo G, Chirico A, Catallo V, Di Lernia D, Riva G (2018) The role of age on multisensory bodily experience: an experimental study with a virtual reality full-body illusion. *Cyberpsychol Behav Social Netw* 21(5):304–310
 154. Ausburn LJ, Martens J, Baukal CE Jr, Agnew I, Dionne RAFB (2019) User characteristics, trait vs. state immersion, and presence in a first-person virtual world. *J Virtual Worlds Res* 12(3)
 155. Tychsen L, Foeller P (2020) Effects of immersive virtual reality headset viewing on young children: visuomotor function, postural stability, and motion sickness. *Am J Ophthal* 209:151–159
 156. Restorick Roberts Amy, Bob De Schutter, Kelley Franks, Elise Radina M (2019) Older adults' experiences with audiovisual virtual reality: perceived usefulness and other factors influencing technology acceptance. *Clin Gerontol* 42(1):27–33
 157. Chiarovano E, Wang W, Rogers SJ, MacDougall HG, Curthoys IS, De Waele C (2017) Balance in virtual reality: effect of age and bilateral vestibular loss. *Front Neurol* 8:5
 158. Schatz R, Regal G, Schwarz S, Suettc S, Kempf M (2018) Assessing the QoE impact of 3D rendering style in the context of VR-based training. In: 2018 10th international conference on quality of multimedia experience. IEEE, pp 1–6
 159. Whitbeck C (1993) Virtual Environments: ethical issues and significant confusions. *Presence Teleoper Virtual Environ* 2(2):147–152
 160. Mel S, Angus A, Adam D, David S, Christoph G, Chris B, Nancy P, Sanchez-Vives MV (2006) A virtual reprise of the Stanley Milgram obedience experiments. *PLoS One* 1(1):e39
 161. Segovia KY, Bailenson JN, Monin B (2009) Morality in Tele-immersive Environments. In: IMMERSCOM, p 17
 162. Williams B, Bernard W (1981) Moral Luck: philosophical papers 1973–1980. Cambridge University Press
 163. Steinicke F, Bruder G (2014) A self-experimentation report about long-term use of fully-immersive technology. In: Proceedings of the 2nd ACM symposium on spatial user interaction, pp 66–69
 164. Ortiz de Gortari AB, Aronsson K, Griffiths M (2011) Game transfer phenomena in video game playing: a qualitative interview study. *Int J Cyber Behav Psychol Learn*. 1(3):15–33
 165. Cantor J (2013) Why horror doesn't die: the enduring and paradoxical effects of frightening entertainment. In: Psychology of entertainment. Routledge, pp 333–346
 166. Lin J-HT (2017) Fear in virtual reality (VR): fear elements, coping reactions, immediate and next-day fright responses toward a survival horror zombie virtual reality game. *Comput Hum Behav* 72:350–361

167. Desurvire H, Kreminski M (2018) Are game design and user research guidelines specific to virtual reality effective in creating a more optimal player experience? Yes, VR PLAY. In: International conference of design, user experience and usability, pp 40–59. Springer
168. Oculus (2021) Oculus, informative guides to help you design, develop, and distribute your VR App. <https://developer.oculus.com/resources/>. Accessed 5 Nov 2021
169. Rolnick A, Lubow RE (1991) Why is the driver rarely motion sick? the role of controllability in motion sickness. *Ergonomics* 34(7):867–879
170. So RHY, Lo WT (1998) Cybersickness with virtual reality training applications: a claustrophobia phenomenon with headmounted displays? In: First world congress on ergonomics for global quality and productivity, Hong Kong
171. Lo WT, So RHY (2001) Cybersickness in the presence of scene rotational movements along different axes. *Appl Ergon* 32(1):1–14
172. Terenzi L, Zaai P (2020) Rotational and translational velocity and acceleration thresholds for the onset of cybersickness in virtual reality. In: AIAA Scitech 2020 forum, p 0171
173. Farmani Y, Teather RJ (2018) Viewpoint snapping to reduce cybersickness in virtual reality. In: Proceedings of the 44th graphics interface conference, pp 168–175
174. Draper MH, Viirre ES, Furness TA, Gawron VJ (2001) Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Hum Factors* 43(1):129–146
175. Davis S, Nesbitt K, Nalivaiko E (2015) Comparing the onset of cybersickness using the oculus rift and two virtual roller coasters. In: Proceedings of the 11th Australasian conference on interactive entertainment (IE 2015), vol 27, p 30
176. Pouke M, Tiiri A, LaValle SM, Ojala T (2018) Effects of visual realism and moving detail on cybersickness. In: 2018 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, pp 665–666
177. Rebenitsch L (2015) Managing cybersickness in virtual reality. *XRDS Crossroads ACM Mag Stud* 22(1):46–51
178. Chardonnet J-R, Mirzaei MA, Merienne F (2021) Influence of navigation parameters on cybersickness in virtual reality. *Virtual Real* 25(3):565–574
179. Steed A, Frlston S, Lopez MM, Drummond J, Pan Y, Swapp D (2016) An in the wild experiment on presence and embodiment using consumer virtual reality equipment. *IEEE Trans Vis Comput Graph* 22(4):1406–1414
180. Parger M, Mueller JH, Schmalstieg D, Steinberger M (2018) Human upper-body inverse kinematics for increased embodiment in consumer-grade virtual reality. In: Proceedings of the 24th ACM symposium on virtual reality software and technology, pp 1–10
181. Choi C, Jun J, Heo J, Kim K (2019) Effects of virtual-avatar motion-synchrony levels on full-body interaction. In: Proceedings of the 34th ACM/SIGAPP symposium on applied computing, pp 701–708
182. Lin JJW, Abi-Rached H, Lahav M (2004) Virtual guiding avatar: an effective procedure to reduce simulator sickness in virtual environments. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 719–726
183. Erickson A, Kim K, Bruder G, Welch GF (2020) Effects of dark mode graphics on visual acuity and fatigue with virtual reality head-mounted displays. In: 2020 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, pp 434–442
184. Jaeger BK, Mourant RR (2001) Comparison of simulator sickness using static and dynamic walking simulators. In: Proceedings of the human factors and ergonomics society annual meeting, vol 45(27), pp 1896–1900. SAGE Publications Sage CA, Los Angeles, CA
185. Hettinger LJ, Riccio GE (1992) Visually induced motion sickness in virtual environments. *Presence Teleoper Virtual Environ* 1(3):306–310
186. Carnegie K, Rhee T (2015) Reducing visual discomfort with HMDs using dynamic depth of field. *IEEE Comput Graph Appl* 35(5):34–41
187. Education Department of Health et al (2014) The Belmont report. Ethical principles and guidelines for the protection of human subjects of research. *J Am Coll Dent* 81(3):4
188. ITU-T Work Item P.IntVR (SG 12) (2021) Subjective test methods for interactive virtual reality applications. https://www.itu.int/itu-t/workprog/wp_item.aspx?isn=17045, Accessed 11 May 2021
189. Brooke J et al (1996) Sus-a quick and dirty usability scale. *Usability Eval Ind* 189(194):4–7
190. Morris JD (1995) Observations: Sam: the self-assessment manikin; an efficient cross-cultural measurement of emotional response. *J Advert Res* 35(6):63–68
191. Poels K, de Kort YAW, IJsselstein WA (2007) D3. 3: game experience questionnaire: development of a self-report measure to assess the psychological impact of digital games
192. IJsselstein WA, De Kort YAW, Poels K (2013) The game experience questionnaire. Technology University of Eindhoven, Eindhoven, pp 3–9
193. Abeele VV, Nacke LE, Mekler ED, Johnson D (2016) Design and preliminary validation of the player experience inventory. In: Proceedings of the 2016 annual symposium on computer–human interaction in play companion extended abstracts, pp 335–341
194. Abeele VV, Spiel K, Nacke L, Johnson D, Gerling K (2020) Development and validation of the player experience inventory: a scale to measure player experiences at the level of functional and psychosocial consequences. *Int J Hum Comput Stud* 135:102370
195. Kourtesis P, Collina S, Doumas LAA, MacPherson SE (2019) Validation of the virtual reality neuroscience questionnaire: maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology. *Front Human Neurosci* 13:417
196. Tcha-Tokey K, Christmann O, Loup-Escande E, Richir S (2016) Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *Int J Virtual Real* 16:33–48. <https://doi.org/10.20870/IJVR.2016.16.1.2880>
197. Bowman DA, Gabbard JL, Hix D (2002) A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence Teleoper Virtual Environ* 11(4):404–424
198. Regal G, Schatz R, Schrammel J, Suetterle S (2018) VRate: a Unity3D asset for integrating subjective assessment questionnaires in virtual environments. In: 2018 10th international conference on quality of multimedia experience, pp 1–3. IEEE
199. Timmerer C, Ebrahimi T, Pereira F (2015) Toward a new assessment of quality. *Computer* 48(3):108–110
200. Skorin-Kapov L, Varela M, Hoffeld T, Chen K-T (2018) A survey of emerging concepts and challenges for QoE management of multimedia services. *ACM Trans Multim Comput Commun Appl* 14(2s):29
201. Gardlo B, Egger S, Hossfeld T (2015) Do scale-design and training matter for video QoE assessments through crowdsourcing? In: Proceedings of the 4th international workshop on crowdsourcing for multimedia. ACM, pp 15–20
202. Clark LA, Watson D (1995) Constructing validity: basic issues in objective scale development. *Psychol Assess* 7(3):309
203. Wilson GM, Angela SM (2000) Do users always know what's good for them? Utilising physiological responses to assess media quality. In: People and computers XIV—usability or else!. Springer, pp 327–339
204. Bouchard S, St-Jacques J, Robillard G, Renaud P (2008) Anxiety increases the feeling of presence in virtual reality. *Presence Teleoper Virtual Environ* 17(4):376–391

205. Murphy D, Higgins C (2019) Secondary inputs for measuring user engagement in immersive VR education environments. arXiv preprint [arXiv:1910.01586](https://arxiv.org/abs/1910.01586)
206. Sheridan TB (1996) Further musings on the psychophysics of presence. *Presence Teleoper Virtual Environ* 5(2):241–246
207. Bailenson JN, Aharoni E, Beall AC, Guadagno RE, Dimov A, Blascovich J (2004) Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. In: *Proceedings of the 7th annual international workshop on presence*, pp 1864–1105
208. Sebastian M, Alexander R (2014) *Quality of experience: advanced concepts, applications and methods*. Springer
209. Kim T, Biocca F (1997) Telepresence via television: two dimensions of telepresence may have different connections to memory and persuasion. *J Comput Med Commun* 3(2):JCMC325
210. Witmer BG, Jerome CJ, Singer MJ (2005) The factor structure of the presence questionnaire. *Presence Teleoper Virtual Environ* 14(3):298–312
211. Schubert TW, Friedmann F, Regenbrecht HT (1999) Decomposing the sense of presence: factor analytic insights. In: *2nd international workshop on presence*, vol 1999
212. Baños RM, Botella C, Garcia-Palacios A, Villa H, Perpiñá C, Alcaniz M (2000) Presence and reality judgment in virtual environments: a unitary construct? *Cyber Psychol Behav* 3(3):327–335
213. Lessiter J, Freeman J, Keogh E, Davidoff J (2000) Development of a new cross-media presence questionnaire: the ITC-sense of presence inventory. Goldsmiths College/Independent Television Commission (UK). Accessed 10 Mar 2003
214. Lombard M, Ditton TB, Crane D, Davis B, Gil-Egüi G, Horvath K, Rossman J, Park S (2000) Measuring presence: a literature-based approach to the development of a standardized paper-and-pencil instrument. In: *3rd international workshop on presence*, Delft, The Netherlands, vol 240, pp 2–4
215. Slater M, McCarthy J, Maringelli F (1998) The influence of body movement on subjective presence in virtual environments. *Hum Factors* 40(3):469–477
216. Usoh M, Arthur K, Whitton MC, Bastos R, Steed A, Slater M, Brooks FP Jr (1999) Walking > walking-in-place > flying, in virtual environments. In: *Proceedings of the 26th annual conference on computer graphics and interactive techniques*, pp 359–364
217. Sas C, O'Hare GMP (2003) Presence equation: an investigation into cognitive factors underlying presence. *Presence Teleoper Virtual Environ* 12(5):523–537
218. Vorderer P, Wirth W, Gouveia FR, Biocca F, Saari T, Jäncke L, Böcking S, Schramm H, Gysbers A, Hartmann T et al (2004) MEC Spatial Presence Questionnaire. Accessed 18 Sept 2015
219. Jennett C, Cox AL, Cairns P, Dhoparee S, Epps A, Tijs T, Walton A (2008) Measuring and defining the experience of immersion in games. *Int J Hum Comput Stud* 66(9):641–661
220. Cheng M-T, She H-C, Annetta Leonard A (2015) Game immersion experience: its hierarchical structure and impact on game-based science learning. *J Comput Assist Learn* 31(3):232–253
221. Makransky G, Lilleholt L, Aaby A (2017) Development and validation of the multimodal presence scale for virtual reality environments: a confirmatory factor analysis and item response theory approach. *Comput Human Behav* 72:276–285
222. Khenak N, Vézien J-M, Bourdot P (2019) The construction and validation of the sp-ie questionnaire: an instrument for measuring spatial presence in immersive environments. In: *International conference on virtual reality and augmented reality*, pp 201–225. Springer
223. Agrawal S, Bech S, Bærentsen K, De Moor K, Forchhammer S (2021) Method for subjective assessment of immersion in audiovisual experiences. *J Audio Eng Soc* 69(9):656–671
224. Mel Slater (2004) How colorful was your day? why questionnaires cannot assess presence in virtual environments. *Presence Teleoper Virtual Environ* 13(4):484–493
225. Slater M, Steed A (2000) A virtual presence counter. *Presence* 9(5):413–434
226. Meehan M, Insko B, Whitton M, Brooks FP Jr (2002) Physiological measures of presence in stressful virtual environments. *ACM Trans Graph* 21(3):645–652
227. Prothero JD, Parker DE, Furness T, Wells M (1995) Towards a robust, quantitative measure for presence. In: *Proceedings of the conference on experimental analysis and measurement of situation awareness*, pp 359–366
228. Slater M, Usoh M, Chrysanthou Y (1995) The influence of dynamic shadows on presence in immersive virtual environments. *Virtual Environ* 95:8–21
229. Lepecq J-C, Bringoux L, Pergandi J-M, Coyle T, Mestre D (2009) Afforded actions as a behavioral assessment of physical presence in virtual environments. *Virtual Reality* 13(3):141–151
230. Blascovich J, Loomis J, Beall AC, Swinth KR, Hoyt CL, Bailenson JN (2002) Immersive virtual environment technology as a methodological tool for social psychology. *Psychol Inquiry* 13(2):103–124
231. Blascovich J, Mendes WB, Hunter SB, Salomon K (1999) Social facilitation as challenge and threat. *J Personal Social Psychol* 77(1):68
232. Strojny PM, Dużmańska-Misiarczyk N, Lipp N, Strojny A (2020) Moderators of social facilitation effect in virtual reality: co-presence and realism of virtual agents. *Front Psychol* 11:1252
233. Bailenson JN, Blascovich J, Beall AC, Loomis JM (2001) Equilibrium theory revisited: mutual gaze and personal space in virtual environments. *Presence Teleoper Virtual Environ* 10(6):583–598
234. Iachini T, Coello Y, Frassinetti F, Ruggiero G (2014) Body space in social interactions: a comparison of reaching and comfort distance in immersive virtual reality. *PLoS One* 9(11):e111511
235. Seele S, Misztal S, Buhler H, Herpers R, Schild J (2017) Here's looking at you anyway! How important is realistic gaze behavior in co-located social virtual reality games? In: *Proceedings of the annual symposium on computer-human interaction in play*, pp 531–540
236. Syrjämäki AH, Isokoski P, Surakka V, Pasanen TP, Hietanen JK (2020) Eye contact in virtual reality—a psychophysiological study. *Comput Human Behav* 112:106454
237. Meehan M, Razaque S, Insko B, Whitton M, Brooks FP (2005) Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. *Appl Psychophysiol Biofeedback* 30(3):239–258
238. Baka E, Stavroulia KE, Magnenat-Thalmann N, Lanitis A (2018) An EEG-based evaluation for comparing the sense of presence between virtual and physical environments. In: *Proceedings of computer graphics international 2018*, pp 107–116. Association for Computing Machinery
239. Baumgartner T, Speck D, Wettstein D, Masnari O, Beeli G, Jäncke L (2008) Feeling present in arousing virtual reality worlds: prefrontal brain regions differentially orchestrate presence experience in adults and children. *Front Human Neurosci* 2:8
240. Kober SE, Kurzmann J, Neuper C (2012) Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: an EEG study. *Int J Psychophysiol* 83(3):365–374
241. Kellogg RS, Kennedy RS, Graybiel A (1964) Motion sickness symptomatology of labyrinthine defective and normal subjects during zero gravity maneuvers. Technical report, Aerospace Medical Research Labs Wright-Patterson AFB Ohio
242. Stone III WB (2017) Psychometric evaluation of the simulator sickness questionnaire as a measure of cybersickness. PhD thesis, Iowa State University

243. Kim HK, Park J, Choi Y, Choe M (2018) Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl Ergon* 69:66–73
244. Golding JF (1998) Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain Res Bull* 47(5):507–516
245. Golding JF (2006) Predicting individual differences in motion sickness susceptibility by questionnaire. *Personal Individ Differ* 41(2):237–248
246. Reason JT (1968) Relations between motion sickness susceptibility, the spiral after-effect and loudness estimation. *Br J Psychol* 59(4):385–393
247. Cebeci B, Celikcan U, Capin TK (2019) A comprehensive study of the affective and physiological responses induced by dynamic virtual reality environments. *Comput Anim Virtual Worlds* 30(3–4):e1893
248. Rebenitsch L, Owen C (2016) Review on cybersickness in applications and visual displays. *Virtual Real* 20(2):101–125
249. Borg G (1998) Borg's perceived exertion and pain scales. *Human Kinetics*
250. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv Psychol* 52:139–183
251. Harris D, Wilson M, Vine S (2020) Development and validation of a simulation workload measure: the simulation task load index (sim-tlx). *Virtual Reality* 24(4):557–566
252. Kim YY, Kim HJ, Kim EN, Ko HD, Kim HT (2005) Characteristic changes in the physiological components of cybersickness. *Psychophysiol* 42(5):616–625
253. Dennison MS, Zachary WA, D'Zmura M (2016) Use of physiological signals to predict cybersickness. *Displays* 44:42–52
254. Wu J, Zhou Q, Li J, Kong X, Xiao Y (2020) Inhibition-related N2 and P3: indicators of visually induced motion sickness (VIMS). *Int J Ind Ergon* 78:102981
255. National Research Council (1981) Procedures for testing color vision: report of working group 41
256. Snellen H (1873) Probabuchstaben zur bestimmung der sehschärfe. H. Peters
257. Fonda G, Anderson M (1988) Fonda-Anderson reading chart for normal and low vision. *Ann Ophthalmol* 20(4):136–139
258. Neely JC (1956) The RAF near-point rule. *Br J Ophthalmol* 40(10):636
259. Imaoka Y, Flury A, de Bruin ED (2020) Assessing saccadic eye movements with head-mounted display virtual reality technology. *Front Psychiatry* 11:922
260. Deary IJ, Liewald D, Nissan J (2011) A free, easy-to-use, computer-based simple and four-choice reaction time programme: the Deary–Liewald reaction time task. *Behav Res Methods* 43(1):258–268
261. Sandberg MA (2011) Cambridge neuropsychological testing automated battery. *Encyclopedia of clinical neuropsychology*. Springer, pp 480–482
262. Pohlmeier AE, Hecht M, Blessing L (2009) User Experience Lifecycle Model ContinUE [Continuous User Experience]. *Der Mensch im Mittep. techn. Syst. Fortschr.-Berichte VDI Reihe* 22:314–317
263. Roto V, Law E, Vermeeren APOS, Hoonhout J (2011) User experience white paper: bringing clarity to the concept of user experience. In: *Dagstuhl seminar on demarcating user experience*, p 12
264. Weiss B, Guse D, Möller S, Raake A, Borowiak A, Reiter U (2014) Temporal development of quality of experience. In: *Quality of experience*. Springer, pp 133–147
265. Seferidis V, Ghanbari M, Pearson DE (1992) Forgiveness effect in subjective assessment of packet video. *Electron Lett* 28(21):2013–2014
266. Murata A, Miyoshi T (2000) Effects of duration of immersion in a virtual environment on postural stability. In: *Smc 2000 Conference Proceedings of 2000 IEEE international conference on systems, man and cybernetics. 'Cybernetics evolving to systems, humans, organizations and their complex interactions' (cat. no. 0, vol 2. IEEE, pp 961–966*
267. Kiryu T, Uchiyama E, Jimbo M, Iijima A (2007) Time-varying factors model with different time-scales for studying cybersickness. In: *International conference on virtual reality*, pp 262–269. Springer
268. Karapanos E, Martens J-B, Hassenzahl M (2009) Reconstructing experiences through sketching. *arXiv preprint arXiv:0912.5343*
269. Wilson G, McGill M (2018) Violent video games in virtual reality: re-evaluating the impact and rating of interactive experiences. In: *Proceedings of the 2018 annual symposium on computer–human interaction in play*, pp 535–548
270. von Wilamowitz-Moellendorff M, Hassenzahl M, Platz A (2006) Dynamics of user experience: how the perceived quality of mobile phones changes over time. In: *User experience-towards a unified view, Workshop at the 4th Nordic conference on human-computer interaction*, pp 74–78
271. Fenko A, Schifferstein HNJ, Hekkert P (2010) Shifts in sensory dominance between various stages of user-product interactions. *Appl Ergon* 41(1):34–40
272. Bailenson JN, Yee N (2006) A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence Teleoper Virtual Environ* 15(6):699–716
273. Klatzky RL, Loomis JM, Beall AC, Chance SS, Golledge RG (1998) Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychol Sci* 9(4):293–298
274. Renner RS, Velichkovsky BM, Helmer JR (2013) The perception of egocentric distances in virtual environments—a review. *ACM Comput Surv* 46(2):1–40
275. Kelly JW, Cherep LA, Siegel ZD (2017) Perceived space in the HTC Vive. *ACM Trans Appl Percept* 15(1):1–16
276. Creem-Regehr SH, Stefanucci JK, Thompson WB, Nash N, McCardell M (2015) Egocentric distance perception in the Oculus Rift (DK2). In: *Proceedings of the ACM SIGGRAPH symposium on applied perception*, pp 47–50
277. Bockelman P, Lingum D (2017) Factors of cybersickness. In: *International conference on human–computer interaction*, pp 3–8. Springer
278. Leow F-T, Ch'ng E, Zhang T, Cai S, See S (2017) In-the-wild observation and evaluation of a Chinese Heritage VR environment with HTC VIVE. In: *International conference on virtual systems and multimedia*, 31 Oct–2 Nov
279. Moustafa F, Steed A (2018) A longitudinal study of small group interaction in social virtual reality. In: *Proceedings of the 24th ACM symposium on virtual reality software and technology*, pp 1–10
280. Antons J-N, Arndt S, Schleicher R, Möller S (2014) Brain activity correlates of quality of experience. In: *Quality of experience*. Springer, pp 109–119
281. Allen RC, Singer MJ, McDonald DP, Cotton JE (2000) Age differences in a virtual reality entertainment environment: a field study. In: *Proceedings of the human factors and ergonomics society annual meeting*, 44(5):542–545. SAGE Publ. Sage CA: Los Angeles, CA
282. Southgate E, Smith SP, Cividino C, Saxby S, Kilham J, Eather G, Scevak J, Summerville D, Buchanan R, Bergin C (2019) Embedding immersive virtual reality in classrooms: ethical, organisational and educational lessons in bridging research and practice. *Int J Child-Comput Interact* 19:19–29
283. Mai C, Wiltzius T, Alt F, Hußmann H (2018) Feeling alone in public: investigating the influence of spatial layout on users' VR

- experience. In: Proceedings of the 10th Nordic conference on human-computer interaction, pp 286–298
284. Eghbali P, Väinänen K, Jokela T (2019) Social acceptability of virtual reality in public spaces: experiential factors and design recommendations. In: Proceedings of the 18th international conference on mobile and ubiquitous multimedia, pp 1–11
285. Saffo D, Yildirim C, Di Bartolomeo S, Dunne C (2020) Crowdsourcing virtual reality experiments using VRChat. In: Extended abstracts of the 2020 chi conference on human factors in computing systems, pp 1–8

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.