



# Assessing User Experience in Virtual Reality – A Comparison of Different Measurements

Carolin Wienrich<sup>1</sup>(✉), Nina Döllinger<sup>2</sup>, Simon Kock<sup>2</sup>,  
Kristina Schindler<sup>2</sup>, and Ole Traupe<sup>2</sup>

<sup>1</sup> University of Würzburg, Würzburg, Germany  
carolin.wienrich@uni-wuerzburg.de

<sup>2</sup> Technische Universität Berlin, Berlin, Germany

**Abstract.** As other interactive technologies, the promising virtual reality (VR) applications and their success highly depend on the quality of the user's experience. The present study gives first insights into relations between general and VR specific aspects of user experience by: (1) Analyzing the evaluation requirements for a large-scale multi-user use case; (2) relating evaluation concepts from the fields of (2D) user experience (UX) and (3D) VR experiences; (3) testing these relations by incorporating measurements from different research fields, and (4) discussing implications for a holistic evaluation framework. During and after experiencing a multi-user adventure on the Immersive Deck of Illusion Walk, participants rated their experience with respect to various components of general UX as well as other components specific to VR experiences. The results revealed positive correlations of presence and social presence with most of the employed post-experience UX measures. The relations between the post- and in-experience measurements showed some inconsistencies. Overall, the experience was positively appraised. The results encourage further investigations into integrating measurements from different lines of research in order to explore the evaluation space of VR experiences.

**Keywords:** User experience · Virtual reality · Evaluation framework

## 1 Introduction

Large-scale virtual reality (VR) experiences have gained importance for location-based entertainment [1]. Companies like Zero Latency [2], the Void [3], or the IMAX Experience Center [4] offer immersive multi-user adventures, which easily supersede the capabilities of typical home VR installations. As with other interactive technologies, promising VR applications and their success highly depend on the quality of the user's experience. While the need for user evaluation is evident, the evaluation procedure itself seems not to be as trivial. The present study performs a user evaluation of a multi-user adventure on the Immersive Deck of Berlin-based Illusion Walk (the first commercial large-scale VR provider in Germany; [5, 6]) by:

- (1) Analyzing the evaluation requirements for a large-scale multi-user use case;
- (2) relating evaluation concepts from the fields of (2D) user experience (UX) and (3D) VR

experiences; (3) testing these relations by incorporating measurements from different research fields, and (4) discussing implications towards a holistic evaluation framework.

## 2 Analyzing the Evaluation Requirements

### 2.1 Situation

The Immersive Deck (Illusion Walk, KG) is a multi-user, multi-room VR installation, which is equipped with a marker-based inside-out tracking technology that allows for the continuous transition between adjacent rooms [5]. Enabling free locomotion, users wear Oculus Rift (Oculus VR, LLC) VR headsets powered by untethered backpack PCs (Intel Core i7 quad-core CPU, NVidia GTX 1070 GPU). In general, Illusion Walk invites users to experience joint adventures, or to collaboratively solve tasks. The basic theme of the story investigated here is a group repair job of a wind turbine in order to restore power to a couple of laser cannons sought to defend Earth's energy supplies against the attack by an alien species. The story requires the participants to move through the VR installation on a predetermined path. Users are virtually represented by avatars based on absolute (6 degrees of freedom; custom built [6]) head and relative (to the head) hand tracking, being rendered from the knees upwards. Hand tracking is achieved using Leap Motion (Leap Motion, Inc.) sensors mounted to the front plates of the VR headsets. Audio stimulation is provided via digital stereo headsets, voice communication is established by means of a custom TeamSpeak server (TeamSpeak Systems GmbH). The virtual environment contains several mixed-reality elements (MRE), in which properties of the real (shape, material, vibration) and the virtual (appearance, sound) world coincide (the walls, a door, a push-button, etc.).

### 2.2 Evaluation Requirements

Assessing the seemingly trivial perception - similarly desired by VR providers and users - of experiencing and mastering an enjoyable challenge in interaction with others involves the evaluation of various concepts from different research fields on multiple levels of complexity.

#### 2.2.1 Aspects of Evaluation

In the field of *human-computer interaction* this task would be addressed by the evaluation of UX. UX concerns itself with the perception and behavior while interacting with products or technical systems (ISO-Norm BS EN ISO 9241-210). When characterizing UX, the functionality, the content, and the aesthetics of a product, the context of use, and the user's perception of and emotions towards the product must be considered [4, 7–9]. Among others, the following researchers proposed dedicated UX models. The *pragmatic/hedonic model* of UX by Hassenzahl (2007), for example, differentiates between the objective qualities of a product and the user's perception of those [10]. Following this model, pragmatic qualities relate to the degree of how the product enables goal achievement, i.e., to usability and usefulness, while the hedonic quality relates to the psychological needs and the emotional experience of the user.

Similar to Hassenzahl's pragmatic quality, the instrumental quality of the *Components of User Experience* (CUE, [9]) model addresses usability and usefulness. The hedonic quality of the CUE model, however, is divided into different modules: non-instrumental product perception (visual aesthetics, commitment, and status), emotions, and consequences of use (product loyalty, intention to use). Another perspective on UX is described by the ISO 6385, which defines *ergonomics* as the understanding of interactions between humans and other elements of a system in order to optimize human well-being and overall system performance [11]. Three fundamental pillars of ergonomics are described, namely physical, cognitive, and organizational ergonomics. *Physical ergonomics* cover the biomechanical design aspects of the equipment used, while *cognitive ergonomics* address mental processes as they affect interactions among elements of a system. *Organizational ergonomics* concern the optimization of socio-technical systems including communication, teamwork, and cooperative work, among others [11].

As important as for other interactive technologies the described UX components are important, but not sufficient concerning the evaluation of the use case (large-scale, multi-user VR). Thus, evaluation concepts often used in VR contexts might be meaningful additions (e.g., [12, 13]).

**VR specific** UX, among other aspects, addresses navigation, wayfinding, and object manipulation (detailed discussion in [12]). Additionally, the degree of user engagement (i.e., presence) and the occurrence and severity of simulator sickness are crucial factors (e.g., [12, 14]). The core concept of *presence* describes the user's feelings of actually being in the place provided by VR (place illusion) and reliably and effectively performing certain actions (plausibility) [15]. According to Schubert et al., the construct of presence has three main components: realism, involvement, and spatial presence [16]. Realism is defined as the user's evaluation of how convincing the virtual environment is. Involvement is defined as a facet of presence based on attention. Spatial presence is defined as a component of spatial construction, i.e., spatial encompassment.

Kinetosis is generally described as a physiological reaction to actual or apparent motion, and includes manifestations such as *simulator sickness* [17]. Simulator sickness usually occurs when motion is presented on a screen introducing substantial visual flow (e.g., simulator: [18]). Apart from common symptoms like discomfort, drowsiness, vomiting, and nausea, simulator sickness can cause further visual and visuo-motoric symptoms like eye strain and dizziness. However, the term is also used for similar symptoms of kinetosis in VR (also referred to as cyber sickness) and other negative side effects of VR experiences [19, 20].

A multi-user context imposes additional requirements. The motive of relatedness from the Self-Determination Theory (SDT: [21]), e.g., addresses the meaning of others for one's own actions as well as the importance of one's own actions for others. Many other factors inducing mutual importance are known from a long tradition of social psychology research (e.g., social identity, social interdependence), a detailed introduction to which is beyond the scope of this article. Some of these aspects, however, are incorporated into the concept of *social presence*. It is defined as the sense of being together in a multi-user VR and covers psychological and behavioral involvement as well as affective aspects [22].

### 2.2.2 Levels of Evaluation

In the field of human-computer interaction evaluation procedures could assess the experience with a whole system (e.g., enjoying the interaction with a smartphone), with individual sub-modules (e.g., enjoying the interaction with a specific app) or with single interactive elements (e.g., enjoying the interaction with the touch display), respectively [23]. Similarly, VR experiences can be evaluated in their entirety (e.g., enjoying an adventure), on a task (e.g., enjoying the mastering of a quest) or element (e.g., enjoying the haptic feeling of mixed reality elements) level. While post-experience questionnaires mainly assess the experience on the system level, observations and in-experience questionnaires can target the experience on system, task or element level, respectively. In the context of a multi-user VR evaluation, in-experience assessments could lead to breaks in presence or story telling. Thus, they cannot be too lengthy and should have plausible ties to the storyline. These requirements raise the question whether (short and adapted to the story) in-experience measurements are valid to assess UX.

The present study investigated general aspects of UX as well as VR specific aspects during and after (i.e., on different levels of) the VR experience described above. Post-experience UX assessment was employed based on the CUE model. In-experience UX was assessed via physical, cognitive, and organizational ergonomics and by assessing affective states. VR specific aspects were assessed through presence, social presence, and health-related issues (e.g., simulator sickness). The study aimed at exploring the relations between these aspects and levels.

## 3 Relating Evaluation Concepts

The approach of Stanney et al. [16] has already been emphasized to include VR specific constructs for evaluating virtual experiences. However, the main components of UX have rarely been related to additional concepts covering the specifics of VR experiences [14]. Furthermore, connections between different levels (e.g., post- and in-experiences measurements) have previously been neglected. The present article proposes the following relations:

Stanney et al. [12] explicitly include presence as one main factor creating compelling VR experiences [20]. Thus, presence should be related to the UX of the investigated application. Concerning hedonic qualities (particularly affective states), affective responses have been shown to heighten the sense of presence in VR and vice versa (e.g., [24, 25]). Hence, presence should particularly be related to the affective measurements incorporated in UX evaluations (post- and in-experience). Furthermore, the study explores how presence is related to the pragmatic quality (i.e., usability and usefulness), non-instrumental aspects (e.g., aesthetic), and the consequences of use, as well as to mental and physical ergonomics (in-experience).

Previously, it was shown that *social presence* positively impacts on game experience [26], on virtual team performance [27], and on the interaction with virtual agents and avatars in VR [28]. However, a direct connection to the above described aspects of UX has not been drawn yet. Therefore, the present study also explores how social presence is connected to *pragmatic* and *hedonic* (particularly affective states) qualities of UX. In addition, it explores how measurements are related to each other among

different levels, particularly to social aspects like the organizational ergonomics (i.e., post- and in-experience).

With respect to VR experiences, *physical ergonomics* might address *pressure points* of the headset or the backpack, as well as distraction by cables and other parts of the equipment limiting the movement of the users. Health-related parameters such as *simulator sickness* might also be important indicators of the specific physical ergonomics of VR systems. As *physical load* (in-experience) is a typical indicator of physical ergonomics (the lack thereof, i.e.), it should be positively related to these VR-specific concepts.

Furthermore, the described in-experience (state affect and ergonomics) and post-experience (CUE model) UX concepts should be tightly connected, as the user experience of individual aspects of the VR experience should clearly affect its overall evaluation. The assessment of affective states during the experience should be related to the emotion module of the CUE model. Mental workload (in-experience), as a typical indicator of cognitive ergonomics, should similarly be related to corresponding post-experience aspects of UX such as negative affects. Organizational ergonomics and physical ergonomics were not connected with post-experience UX.

In sum, the study addressed three goals:

- (1) Assessing the assumed relations between the described evaluation concepts originating from the fields of general and VR specific UX.
- (2) Exploring relations between the concepts described on different levels (post- and in-experience).
- (3) Detecting overlaps, and therefore potential redundancies, in order to work towards an integrated evaluation framework for (large-scale, multi-user) VR applications.

## 4 Empirical Testing of Relations

The present study puts different evaluation concepts and potential relations between them to the test by evaluating a beta version of the first large-scale, multi-user VR experience of Illusion Walk in Germany [5, 6]. The research questions are explorative. It was assumed that different concepts from different fields of research (general and VR-specific UX) assessed on different levels (post- and in-experience) are related to each other. Correlative analyses examined these relations and proposed conclusions for an evaluation framework of large-scale multi-user VR applications.

Note that the present tests were part of a bigger experimental cycle with additional research questions. To answer these, two experimental conditions were established. In the interdependence condition, (IDP) participants had to solve a series of tasks together with their fellow participants while mutually depending on each other's performance. In the non-interdependence condition (nIDP) participants had to solve a similar control task on their own. In line with the expectations, stronger team affiliation and more cooperation (i.e., mutual importance) were found for participants in the IPD versus the nIPD condition (results presented in detail in [29]).

## 4.1 Method

### 4.1.1 Participants

Seventy-two volunteers ( $n = 12$  female; mean age 32.11 years;  $SD = 8.68$  years) with normal or corrected-to-normal vision participated in the study. Participants conducted the experiment in groups of three ( $n = 4$  female experimental groups). None of the participants reported any health problems such as epilepsy or migraine (which could be triggered by VR). A screening questionnaire revealed that the sample was rather highly experienced with VR ( $M = 3.35$ ,  $SD = 1.49$ ; poles of scale 1 to 5) and video gaming ( $M = 3.68$ ,  $SD = 1.20$ ; poles of scale 1 to 5). Participants also reported a high technical affinity ( $M = 4.42$ ;  $SD = .78$ ; poles of scale 1 to 5) and a good tolerance for simulator sickness ( $M = 4.32$ ;  $SD = .82$ ; poles of scale 1 to 6). As compensation, participants received a voucher from Illusion Walk for a free VR experience.

### 4.1.2 Materials

**VR Installation and Equipment.** The experiment was conducted in the multi-user, multi-room VR installation “Immersive Deck” (Illusion Walk, KG), which is equipped with a marker-based inside-out tracking technology that allows for the continuous transition between adjacent rooms (see Subsect. 2.1). The virtual environment was created and presented with Unity3D (Unity Technologies) running a client-server model over 802.11ac Wi-Fi connections.

**Questionnaires.** All text-based material for informing, screening, instructing, and assessing the participants outside the experience was presented on tablet devices. To assess the participants’ state during different parts of the experience, an in-experience questionnaire was set up (see Fig. 1).



**Fig. 1.** In-experience questionnaires presented in a pop-up style and operated via the participants’ tracked hands; the example shows an item of the PANAS (German version).

**Post-experience Assessment.** *UX* was assessed by the modules of the *meCUE* questionnaire (*modular evaluation of key Components of User Experience*, [30]). The modules address *product perception* (instrumental - pragmatic subscales: *usability* and *usefulness*; non-instrumental - hedonic subscales: *visual aesthetics*, *commitment*, and *status*), *emotions* (subscales: *positive affect* and *negative affect* - hedonic), and *consequences of use* (subscales: *product loyalty*, *intention to use*), as well as an *overall evaluation* of the experience (one item). In addition, participants were asked: “What are you willing to pay for a similar experience lasting two hours?”.

The sense of **presence** was measured with the German version of the *iGroup Presence Questionnaire (iPQ)*, [16]) entailing subscales for *general presence*, *realism*, *involvement*, and *spatial presence*.

**Social presence** was measured by the *Social Presence Module* of the *Game Experience Questionnaire (GEQ)*, [31]). It includes *psychological involvement*, *behavioral involvement*, and *negative feelings*.

A **discomfort scale** was used following a model, which assumes that discomfort is influenced by biomechanical design aspects, such as pressure points, and therefore is more relevant to the ergonomic side of design [31]. Since the original discomfort scale was designed to assess seat comfort, the items were slightly adapted to reflect discomfort arising from the VR equipment (i.e., pressure points of the headset, eye strain, etc.).

An itemized analysis is not recommended for the *meCUE*. Nonetheless, specific items (negative affect subscale) were used assess adverse effects of the experience, such as tiredness.

**In-experience Assessment.** Complementing post-experience measures, the in-experience assessment contained one team-based item (organizational ergonomics) concerning the importance of the other group members (“At the moment, the experience with the other experts is important for me.”).

**Physical load** (physical ergonomics) and **mental workload** (mental ergonomics) were similarly measured by one item, which had been adapted from a scale assessing experienced strain (*SEA*, [33]): “At the moment, how physically strained do you feel?”; “At the moment, how mentally strained do you feel?”. Answers could range from “not strained at all” to “extremely strained”.

The participants’ **affective state** was assessed during the experience through the *Positive and Negative Affect Schedule (PANAS)*, [32]).

In addition, participants reported their level of **simulator sickness** on the *Fast Motion Sickness Scale (FMS)*, [33]) before, during, and after the experience.

As mentioned above, the present experiment was part of a bigger experimental cycle with additional research questions. Thus, some further measures were assessed but not referred to here. These include several physiological parameters and behavioral observations, the latter being recorded by a supervisor covertly following the group of participants throughout the tracking space. Furthermore, the *Game Experience Questionnaire (GEQ)*, [31]) was administered. It includes the *In-Game Module* with its subscales *competence*, *flow*, *immersion*, *challenge*, *tension*, *negative* and *positive affect*, and the *Post-Game Module* entailing the subscales *positive* and *negative experiences*, *tiredness*, and *returning to reality*. Moreover, the cooperative module of the *Competitive*

and Cooperative Presence in Gaming Questionnaire (CCPIG, [34]), elaborating on cooperative social presence, was administered to investigate the effects of the social interdependence manipulation. Some additional open questions (post-experience) were also not reported in the present article. All details are published in [29].

### 4.1.3 Procedure

**Overall Structure.** Each session was structured into a preparation, an experience, and a post-experience phase. The *preparation phase* included usage instructions and safety warnings for the Immersive Deck as well as retrieving the participants' informed consent. Demographic and health data were assessed via a screening questionnaire. The *experience phase* started after putting on the VR equipment followed by a brief technical check-up. The VR experience let the participants move through the virtual scene following the requirements created by a predefined set of events (see below), including a series of tasks and in-experience assessments. Behavioral observations were conducted from outside of the virtual environment. The experience phase ended with unmounting the equipment. The *post-experience phase* included the completion of the post-experimental questionnaires together with a series of open questions.

**Detailed Description of the Experience.** The experience began with the collective exploration of the virtual starting room, which resembled the physical starting room of the VR installation (MRE) - a measure to facilitate presence via a gradual transition into the virtual environment. The adventure could then be started by any of the participants by pressing a push-button (MRE). This and other elements of the scene could be operated via the virtually represented hands. Immediately following the button press, the first in-experience assessment (pop-up questionnaire; see Fig. 1) was performed establishing a baseline measurement.

The subsequent storyline was structured into three sections, each of them leading to an instance of the task, followed up by a repetition of the in-experience assessment. The task was performed on a graphical user interface, required visuo-motor skills, and had a pronounced speed component. In one condition it also required coordination between participants, conveying an experimental manipulation (nIDP/IDP) in the context of another research question [29].

The basic theme of the story was a collective repair job of a wind turbine to restore power to a set of laser cannons defending Earth's energy supplies against a hostile alien attack. In short, the participants had to enter the wind turbine facility, repair the laser cannons, and - after being abducted by the aliens - activate a spaceship's self-destruction mechanism to ultimately fend off the attack. The story is outlined in greater detail in [29]. To ensure the participants' motivation, the story seemingly depended on task success, with minor contextual workarounds (not detailed here) allowing it to progress even in the case of a failure. The in-experience questionnaires were disguised as a state evaluation within the initially suggested work context.

**Task Description.** The recurrent task was performed as a minigame linked to a pedestal with identical operating panels on its three side faces (see [29]). The pedestal appeared at pre-designed positions in the scene. Each participant's panel contained



three differently shaped and colored buttons (red square, blue circle, green triangle), a graphical timer for the trial time, a numerical timer for the total task time, and a progress bar of stacked triangles on top of the pedestal, its number representing the current sum of successful (positive) and failed (negative) trials. Across the task instances, the difficulty increased to keep the task interesting and challenging.

## 4.2 Data Analysis

To explore the relations between the different UX measures and components, correlations were computed. In a first step, the items of the standardized questionnaires were aggregated according to the corresponding manuals. Concerning the in-experience questions the average of each scale over the four measurement points were calculated. As most of the resulting scales did not fulfill the requirements for the parametric Pearson correlation coefficient, Spearman's rank correlations were calculated.

## 5 Results

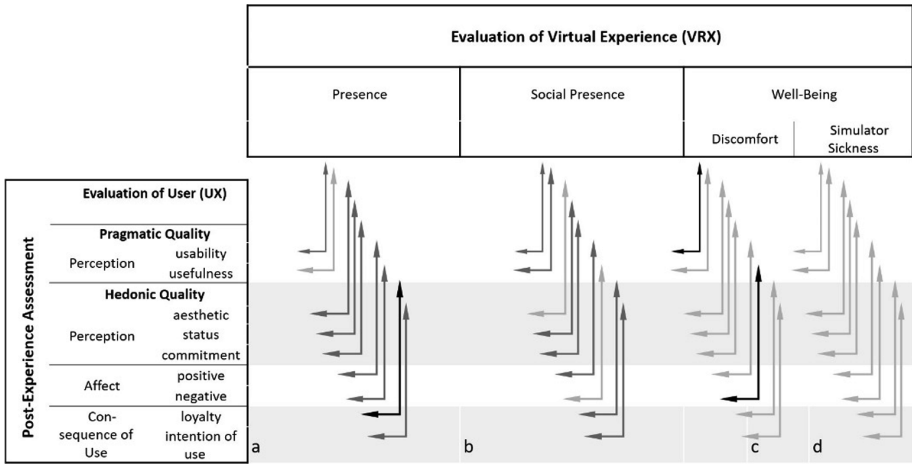
Due to technical problems in some in-experiment questionnaires, the sample size for in-experience analysis was reduced to  $N = 38$  (9 female, 29 male) participants. Thus, the correlation analyses concerning the post-experience measurements are based on 72 participants; those concerning in-experience measurements included 38 participants. Tables 1 and 2 show descriptive statistics as well as the results of correlation analyses between the post-experience measurements. Figure 2 illustrates the respective correlation patterns.

### 5.1 Relations Between Presence (VR Specific) and Post-experience UX

In general, the VR experience of Illusion Walk enabled a strong feeling of *presence* (Table 1; caption row). Overall, presence correlated with all aspects of the post-experience UX evaluation (subscales of meCUE) except *usefulness* (see Fig. 2, panel a) at least in one subscale (*involvement*, *realness*, *spatial presence*, or *general presence*). A closer look revealed the correlation with *usability* and *positive affect* was only significant for the *general presence* value (Table 1; column 4; rows 1 and 6). Particularly, *visual aesthetic* and *loyalty* were correlated with *realness* and *spatial presence* (Table 1; columns 2 and 3; rows 3 and 8).

### 5.2 Relations Between Social Presence (VR Specific) and Post-experience UX

Analogous to the feeling of presence, the VR experience of Illusion Walk enabled a strong feeling of *social presence* (Table 2, caption row). Overall, social presence correlated with all aspects of the post-experience UX evaluation (subscales of meCUE), except *usefulness* and *negative affect* (see Fig. 2, panel b) at least in one subscale (*involvement*, *realness*, *spatial presence*, or *general presence*). Particularly, the



**Fig. 2.** Arrows depict the correlations between post-experience UX (meCUE) and presence, social presence, discomfort, and simulator sickness (VR specific measurements). Black arrows indicate a significant correlation between each subscale of the VR specific measurement and the subscales of the meCUE. Dark grey arrows indicate a significant correlation between at least one subscale of the VR specific measurement and the subscales of the meCUE. Light grey arrows indicate any significant correlation between subscales of the VR specific measurement and the subscales of the meCUE.

subscale *empathy* was correlated with *pragmatic quality*, *positive affect*, *status*, *commitment*, and *consequences of use* (Table 2; column 1; rows 1, 2, 4, 5, 6, 8, 9). Surprisingly, the subscale *negative feelings* showed positive correlations with *status* and *commitment* (Table 2; column 2; rows 4 and 5). The subscale *behavior* did not show any significant correlation with the subscales of the meCUE (Table 2; column 3).

### 5.3 Relations Between Health Measures (VR Specific) and Post-experience UX

In general, the VR installation of Illusion Walk caused very low values of *simulator sickness* ( $M = 1.31$ ;  $SD = .46$ ) and satisfying levels of *discomfort* (Table 2, caption row and Fig. 2, panel c and d).

*Simulator sickness* was not correlated significantly with any subscale of the meCUE questionnaire. Probably, the low variances of its sub-modules are responsible for this observation. Similarly, the *discomfort* scale showed only two significant correlations with the meCUE subscales, namely *usability* (Table 2, column 4; row 1) and *negative affect* (Table 2; column 4; row 7).

### 5.4 Relations Between Post-experience Measures and In-experience UX

In contrast to the correlation patterns reported above, only two correlations between in- and post-experience UX reached significance (*presence* subscales *spatial* and

**Table 1.** Means and standard deviations of the different measures and the correlations between presence (iPQ) and post-experience UX (meCUE).

	<b>Presence</b>	<i>Involv.</i> (4.6; 1.28)	<i>realness</i> (3.16; 1.27)	<i>spatial</i> (5.44; 0.86)	<i>general</i> (5.97; 0.66)
<b>Pragmatic Quality Perception</b>	<i>usability</i> (5.64; 0.9)	$r_s < -.01$ , p = .99	$r_s = .12$ , p = .32	$r_s = .19$ , p = .11	$r_s = .29$ , p = .02
	<i>usefulness</i> (5.64; 0.89)	$r_s = -.10$ , p = .39	$r_s = .13$ , p = .26	$r_s = .08$ , p = .48	$r_s = .10$ , p = .38
<b>Hedonic Quality Perception</b>	<i>visual aesthetic</i>	$r_s = .20$ , p = .10	$r_s = .54$ , p < .01	$r_s = .38$ , p < .01	$r_s = .22$ , p = .07
	<i>status</i> (3.69; 1.23)	$r_s = .05$ , p = .65	$r_s = .33$ , p < .01	$r_s = .07$ , p = .58	$r_s = .23$ , p = .05
	<i>Commitment</i>	$r_s = -.14$ , p = .23	$r_s = .30$ , p < .01	$r_s < .01$ , p = .98	$r_s = .19$ , p = .11
<b>Hedonic Quality Affect</b>	<i>positive</i> (4.62; 0.83)	$r_s = .13$ , p = .27	$r_s = .17$ , p = .14	$r_s = .17$ , p = .16	$r_s = .25$ , p = .04
	<i>negative</i> (2.38; 1.11)	$r_s = -.35$ , p < .01	$r_s = -.14$ , p = .25	$r_s = -.34$ , p < .01	$r_s = -.47$ , p < .01
<b>Cons. of Use</b>	<i>loyalty</i> (3.68; 1.15)	$r_s = .33$ , p = .01	$r_s = .41$ , p < .01	$r_s = .23$ , p = .05	$r_s = .28$ , p = .02
	<i>intention of use</i>	$r_s = .24$ , p = .04	$r_s = .22$ , p = .06	$r_s = .20$ , p = .09	$r_s = .32$ , p < .01

**Table 2.** Means and standard deviations of the different measures and the correlations between social presence (GEQ), well-being (discomfort scale) and post-experience UX (meCUE).

		Social Presence	<i>empathy</i> (3.41; 0.66)	<i>behavior</i> (3.51; 0.83)	<i>feelings</i> (1.94; 0.46)	<i>discomfort</i> (2.37; 1.2)
Pragmatic Quality	Perception	<i>usability</i> (5.64; 0.9)	$r_s = .26,$ $p = .03$	$r_s = .12,$ $p = .33$	$r_s = .08,$ $p = .49$	$r_s = -.28,$ $p = .02$
		<i>usefulness</i> (5.64; 0.89)	$r_s = .28,$ $p = .02$	$r_s = .19,$ $p = .12$	$r_s = .24,$ $p = .05$	$r_s = .48,$ $p = .08$
Hedonic Quality	Perception	<i>visual aesthetic</i>	$r_s = .16,$ $p = .14$	$r_s = .07,$ $p = .54$	$r_s = .06,$ $p = .63$	$r_s = -.21,$ $p = .08$
		<i>status</i> (3.69; 1.23)	$r_s = .36,$ $p = .02$	$r_s = .15,$ $p = .22$	$r_s = .35,$ $p = .01$	$r_s < .01,$ $p = .98$
		<i>commitment</i> (2.24; 1.18)	$r_s = .31,$ $p = .01$	$r_s = .09,$ $p = .46$	$r_s = .24,$ $p = .04$	$r_s = .03,$ $p = .78$
Affect	<i>positive</i> (4.62; 0.83)	$r_s = .27,$ $p = .02$	$r_s = .11,$ $p = .36$	$r_s = .09,$ $p = .45$	$r_s = -.14,$ $p = .24$	
	<i>negative</i> (2.38; 1.11)	$r_s = -.15,$ $p = .19$	$r_s = -.1,$ $p = .41$	$r_s = .08,$ $p = .49$	$r_s = -.04,$ $p = .76$	
Cons. of Use	<i>loyalty</i> (3.68; 1.15)	$r_s = .24,$ $p = .04$	$r_s = .07,$ $p = .64$	$r_s = .22,$ $p = .06$	$r_s = -.03,$ $p = .82$	
	<i>intention of use</i>	$r_s = .36,$ $p = .01$	$r_s = .08,$ $p = .51$	$r_s = .05,$ $p = .23$	$r_s = .04,$ $p = .76$	

*co-presence*:  $r_s = 0.41$ ,  $p = 0.02$ ; and *social presence* subscales *behavior* and *co-presence*:  $r_s = 0.40$ ,  $p = 0.02$ ). However, *physical load* (in-experience) correlated positively with *discomfort* (VR specific post-experience) ( $r_s = .39$ ,  $p = .01$ ). Similarly, a positive correlation between *physical workload* and *simulator sickness* (VR specific) was found ( $r_s = .31$ ,  $p = .04$ ). The expected correlations between the post- and in-experience UX measurements (affect module of meCUE and PANAS; *negative affect* subscale of meCUE and *mental workload*) could also be found.

## 6 Discussing Implications Regarding a Holistic Evaluation Framework

The emerging public perception of VR leads to a growing number of location-based entertainment centers offering multi-user VR experiences (e.g. Zero Latency with currently 12 sites). While the need for user evaluation is evident, the evaluation procedure itself seems not to be as trivial. The present study performs a user evaluation of a multi-user adventure on the Immersive Deck of Berlin's Illusion Walk (the first commercial large-scale VR provider in Germany) [5, 6] by: (1) Analyzing the evaluation requirements for a large-scale multi-user use case; (2) relating evaluation concepts from the fields of (2D) user experience (UX) and (3D) VR experiences; (3) testing these relations by employing measurements from different research fields, and (4) discussing implications for a holistic evaluation framework.

The present study applied UX concepts from the field of human-computer interaction as well as VR specific aspects to appraise the experience of users. The modules of the meCUE (based on the Components of User Experience model, [30]) were related to presence, social presence and health related measurements like simulator sickness (VR specific aspects). The correlation patterns revealed that particularly presence and social presence were related to the components of UX (measured post-experience).

**Presence.** While the association between presence and affect is well documented (e.g., [24, 25]), the relation between presence and the other UX aspects is mostly unexplored. In the research history of presence, the impact of immersion on presence is well-established (e.g., [15]). Immersion is defined as the degree to which a person can be engrossed in a virtual world, based on objective and quantifiable multisensory stimuli. Hence, immersion describes the extent to which the technological features of the device and the setting can provide the user with the illusion of reality. The higher the degree of immersion, the higher the potential feeling of presence (e.g., [35]). On the one hand, the association between pragmatic aspects of UX and presence might merely mirror the relation between immersion and presence. On the other hand, presence occurs when a mental model is constructed, and attention is allocated to a virtual environment [15, 16]. Hence, the association between pragmatic aspects of UX and presence might also reflect the degree of attention that is deployed to inaccuracies of the systems. Further research is necessary to clarify the causal direction of the association and to transfer the findings on the construct of presence to the field of UX.

**Social Presence.** Due to the present multi-user context, mutual importance as well as the sense of being together are crucial evaluation aspects. Previously, it was shown that

social presence positively impacts on game experience [26], on virtual team performance [27], and on the interaction with virtual agents and avatars in VR [28]. Further, it is well known that the presence of others influences cognition and behavior (cf. social cognition, [36]), which has also been shown in VR (e.g., [37]). These previous results might indicate a relation between social presence and UX aspects, but a direct connection to the aspects of UX described above has not been drawn yet. The results of the present study revealed that mainly the subscale of empathy contributed to these relations while particularly the subscale of behavior did not show any relations. However, this does not justify the conclusion that the other subscales are irrelevant for VR user evaluations. Rather, the findings might indicate additional information, which is provided by social presence compared to general UX concepts.

The health-related measurements only sporadically showed relations to the UX aspects. The low values of discomfort and simulator sickness in this study might be able to explain these findings. Average scores near either end of a scale usually exhibit only a limited variance, which in turn can only result in limited correlations. Higher values on health-related measures might have produced stronger correlations with the more traditional UX concepts, even if the underlying causal relations are unaffected. Thus, health-related aspect should still be considered in future evaluation processes.

In sum, the present study revealed relations between aspects of UX and VR specific measures. On the one hand, the results encourage to consider measurements from different lines of research to explore the evaluation space of VR experiences. On the other hand, the results motivate to consolidate the spirit of user-centered design processes in commercial VR contexts. However, the present study is just a first step towards an appropriate evaluation framework. Clearly, more work is needed, which should include the following steps:

**Methods and Theory Building.** As the applied methodology was part of a larger experimental circle, only correlation patterns were calculated to get an impression of assumed relations. However, causal conclusion could not be drawn. Hence, the proposed relations should be examined in future, experimental studies.

In order to examine the contributions of different UX-concepts to the evaluation of the experience of users, factorial analyses would reveal variance shared by different concepts. Such results could indicate measurements addressing the same concepts and therefore being redundant, and separate them from measurements, which cover different concepts and should be used together. Future theories should conflate these findings into a model predicting VR-UX, which future research should experimentally substantiate.

Another question concerns the point of evaluation time. In the context of multi-user VR evaluation, in-experience assessments could lead to breaks in presence or interaction. Hence, the question occurs whether (short and adapted to the story) in-experience measurements are valid to assess UX. Our results revealed a satisfying accordance between the post-experience and in-experience UX measurements. The health-related measurements (VR specific) and the in-experience UX measurements also correlated positively. However, presence and social presence were not related to the in-experience UX measurements. The latter result contrasts with the relation to the post-experience measurements. Particularly the inconsistencies regarding the affective

measurements are surprising. Previous studies showed strong relations between the feeling of presence and affects [24, 25].

**Additional Indicators.** Stanny et al. [12] stated wayfinding, navigation, and object manipulation to be important aspects of VR evaluation. In addition, system parameters, particularly latencies, impacted on the experience of users and hence should be evaluated as well. Another very important step are qualitative analyses. VR experiences are often described as journeys. Hence, evaluating the smoothness of the experience would improve the evaluation framework. Benford et al. [38] suggested the analysis of transitions and trajectories. Trajectories reveal the continuity, coherence, and interaction patterns between users and the equipment, as well as between different users. Continuity deals with various transitions within the VR experience itself, but also between the VR experience and the experience in the real world (e.g., transitions of time, space, or roles). Similarly, (2D) UX testing includes qualitative analyses (e.g., think aloud, observation) which should be considered for an evaluation framework. Another related question concerns the point of evaluation time and therewith the level that can be evaluated. Post-experience measurements often assess the holistic experience. In contrast, in-experience evaluations might stress challenges concerning specific tasks or interactions within the experience. However, the in-experience questions should conflate into the story to avoid breaks in presence or interactions [39]. In our opinion, VR in particular might not only be the object of evaluation, but also provide a versatile tool for such analyses. Many parameters can be recorded and controlled - e.g., the trajectories of avatars can be observed without additional camera equipment and easily compared to ideal trajectories.

In sum, the present study put different evaluation tools (general UX and VR specific) and potential associations between them to the test by evaluating the beta version of the first large-scale multi-user VR experience of Illusion Walk in Germany [5, 6]. The tracking and interaction technology of the Immersive Deck seems to have contributed to a general positive evaluation of the experience. High ratings of presence and social presence and low in-experience negative affect indicate the absence of any major constraints due to the equipped VR hardware: Users experienced an enjoyable challenge in interaction with others. The present paper represents a first step towards integrating evaluation concepts from different research fields in order to evaluate large-scale multi-user VR experiences.

## References

1. The Future of Virtual Reality isn't Your Living Room - It's The Mall. <https://www.forbes.com/sites/charliefink/2017/03/07/the-future-of-virtual-reality-isnt-your-living-room-its-the-mall/1#1e3398502bec>. Accessed 22 Dec 2017
2. Zero Latency Madrid: Google Maps (2017). <https://zerolatencyvr.com/>. Accessed: 22 Dec 2017
3. The void (2007)
4. Understanding, Scoping and Defining User Experience: A survey approach. dl.acm.org. <http://dl.acm.org/citation.cfm?id=1518813>. Accessed 22 Dec 2017

5. Home—Illusion Walk Berlin the Virtual Reality Immersive Experience Provider. <https://www.illusion-walk.com/>. Accessed 22 Dec 2017
6. Podkosova, I., Vasylevska, K., Schoenauer, C., Vonach, E., Fikar, P., Bronederk, E., Kaufmann, H.: ImmersiveDeck: a large-scale wireless VR system for multiple users (2016). <https://ieeexplore.ieee.org>
7. ISO 9241-11: Guidelines for specifying and measuring usability (1993)
8. McCarthy, J., Wright, P.: Technology as experience. *Interactions* **11**(5), 42 (2004)
9. Thüring, M., Mahlke, S.: Usability, aesthetics and emotions in human-technology interaction. *Int. J. Psychol.* **42**(4), 253–264 (2007)
10. Law, E., Vermeeren, A., Hassenzahl, M., Eds, M.B.: Towards a UX manifesto. In: COST294-MAUSE Affiliated Workshop, Structure, September 2007, vol. 2, pp. 205–206 (2007)
11. International Ergonomics Association. <http://www.iea.cc/>. Accessed 22 Dec 2017
12. Stanney, K.M., Mourant, R.R., Kennedy, R.S.: Human factors issues in virtual environments: a review of the literature. *Presence Teleoper. Virtual Environ.* **7**(4), 327–351 (1998)
13. Tromp, J.G., Steed, A., Wilson, J.R.: Systematic usability evaluation and design issues for collaborative virtual environments. *Presence Teleoper. Virtual Environ.* **12**(3), 241–267 (2003)
14. Wienrich, C., Noller, F., Thüring, M.: Design principles for VR interaction models: an empirical pilot study. In: *Virtuelle und erweiterte Realitäten 14. Workshop der GI-Fachgruppe VR/AR*, pp. 162–171 (2017)
15. Slater, M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. Roy. Soc. B Biol. Sci.* **364**(1535), 3549–3557 (2009)
16. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: factor analytic insights. *Presence Teleoper. Virtual Environ.* **10**(3), 266–281 (2001)
17. Hegemann, S.: Die Entstehung von Kinetosen. In: Plinkert, P.K., Klingmann, C. (eds.) *Hören und Gleichgewicht*, pp. 185–193. Springer, Vienna (2010). [https://doi.org/10.1007/978-3-211-99270-8\\_19](https://doi.org/10.1007/978-3-211-99270-8_19)
18. Kennedy, R.S., Drexler, J., Kennedy, R.C.: Research in visually induced motion sickness. *Appl. Ergon.* **41**(4), 494–503 (2010)
19. Burdea, G.C., Coiffet, P.: *Virtual Reality Technology*. Wiley-IEEE Press, New York (2003)
20. Stanney, K.M., Kennedy, R.S., Drexler, J.M.: Cybersickness is not simulator sickness (1997). <http://journals.sagepub.com/>
21. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* **55**(1), 68–78 (2000)
22. Ijsselstein, W., de Kort, Y., Poels, K., Jurgelionis, A., Bellotti, F.: Characterising and measuring user experiences in digital games. In: *ACE Conference 2007, January 2007*
23. Reinhardt, D., Kuge, J., Hurtienne, J.: CHAI: coding heuristics for assessing intuitive interaction. In: Marcus, A., Wang, W. (eds.) *DUXU 2018. LNCS*, vol. 10918, pp. 528–545 (2018)
24. Baños, R., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., Rey, B.: Immersion and emotion: their impact on the sense of presence. *CyberPsychol. Behav.* **7**(6), 734–741 (2004)
25. Riva, G., et al.: Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychol. Behav.* **10**(1), 45–56 (2007)
26. Ryan, R.M., Rigby, C.S., Przybylski, A.: The motivational pull of video games: a self-determination theory approach. *Motiv. Emot.* **30**(4), 347–363 (2006)
27. Hertel, G., Konradt, U., Orlikowski, B.: Managing distance by interdependence: goal setting, task interdependence, and team-based rewards in virtual teams. *Eur. J. Work Organ. Psychol.* **13**(1), 1–28 (2004)



28. Ellis, J.B., Luther, K., Bessiere, K., Kellogg, W.A.: Games for virtual team building. In: Proceedings of the Designing Interactive Systems, DIS 2008, pp. 295–304 (2008)
29. Wienrich, C., Schindler, K., Döllinger, N., Kock, S., Traupe, O.: Social presence and cooperation in a large-scale, multi-user virtual reality – an empirical evaluation of a location-based adventure. In: IEEE VR (2018)
30. Minge, M., Riedel, L.: meCUE – Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In: Mensch und Computer, Munich, pp. 89–98 (2013)
31. IJsselsteijn, W.A., de Kort, Y.A.W., Poels, K.: Game Experience Questionnaire. FUGA the fun of gaming: Measuring the human experience of media enjoyment GAME (2015)
32. Krohne, W.H., Egloff, B., Kohlmann, C.-W., Tausch, A.: Untersuchungen mit einer deutschen Version der “Positive and Negative Affect Schedule” (PANAS). *Diagnostica* **42** (2), 139–156 (1996)
33. Keshavarz, B., Hecht, H.: Validating an efficient method to quantify motion sickness. *Hum. Factors* **53**(4), 415–426 (2011)
34. Hudson, M., Cairns, P.: Measuring social presence in team-based digital games. In: Interacting with Presence HCI Sense Presence Computer Environments, p. 83 (2014)
35. Dörner, R., Broll, W., Grimm, P., Jung, B. (eds.): Virtual und Augmented Reality (VR/AR). Springer, Heidelberg (2013). <https://doi.org/10.1007/978-3-642-28903-3>
36. Skarratt, P.A., Cole, G.G., Kingstone, A.: Social inhibition of return. *Acta Psychol. (Amst)* **134**(1), 48–54 (2010)
37. Wienrich, C., Gross, R., Kretschmer, F., Müller-Plath, G.: Developing and proving a framework for reaction time experiments in VR to objectively measure social interaction with virtual agents. In: IEEE VR (2018)
38. Benford, S., Giannachi, G., Koleva, B., Rodden, T.: From interaction to trajectories: designing coherent journeys through user experiences. In: Chi 2009 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, vol. 1–4, pp. 709–718 (2009)
39. Freytag, S., Wienrich, C.: Evaluation of a virtual gaming environment designed to access emotional reactions while playing. In: Virtual Worlds and Games for Serious Applications (VS-Games), pp. 145–148. IEEE (2017)