



Questionnaires embedded in virtual environments: reliability and positioning of rating scales in virtual environments

Georg Regal¹ · Jan-Niklas Voigt-Antons² · Steven Schmidt² · Johann Schrammel¹ · Tanja Kojić² · Manfred Tscheligi¹ · Sebastian Möller²

Received: 9 April 2019 / Published online: 30 October 2019
© Springer Nature Switzerland AG 2019

Abstract

Current developments in virtual reality (VR) hardware have made immersive VR experiences more affordable through commercially available head-mounted displays. As more studies are likely to be conducted using these devices, the question arises how to embed questionnaires in virtual environments without impairing the immersive user experience. In this work we investigate two different aspects: (1) if a rating performed in a virtual environment is comparable to a rating obtained via a paper questionnaire and (2) how questionnaires for assessing virtual experiences should be designed and integrated into the virtual environment. For this research, we used our own extended version of VRate—a VR questionnaire asset for Unity. In the first study with 27 participants, we compared ratings assessed within VR with ratings obtained using a paper questionnaire. We found that the ratings gathered in VR are comparable to the ratings gathered in the real world by paper-pencil questionnaires (subscales: global presence, spatial presence, and experience realism). In the second study with 48 participants, we investigated the users' perceived suitability of the VR questionnaire and the optimal mounting position of the questionnaire (hand-mounted, head-up display or billboard). Moreover, we investigated whether the questionnaire should be answered in the same or in a separate dedicated virtual environment and how the users' feeling of presence in VR is influenced by this placement. Results indicate a subjective preference for the billboard position, with a significant preference for billboard over hand-mounted and no significant preference between billboard and head-up-display. Regarding the placement of the VR questionnaire (in-scene vs. dedicated virtual environment) we did not find any influence on presence. In the following, we discuss the pros and cons of different placement/mounting options and provide suggestions for designing and implementing questionnaires embedded in virtual environments.

Keywords Virtual reality · Subjective rating · Test method

Georg Regal and Jan-Niklas Voigt-Antons have contributed equally to this publication.

✉ Jan-Niklas Voigt-Antons
jv@qu.tu-berlin.de

Georg Regal
georg.regal@ait.ac.at

¹ AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Wien, Austria

² Quality and Usability Lab, Technische Universität Berlin, Berlin, Germany

Introduction

Innovation in Virtual Reality (VR) has made fully immersive experiences affordable for consumers through commercially available head-mounted displays and input devices (cf. [1]). A broad range of application scenarios for virtual environments (VE) exists and is steadily evolving.

Based on these developments, immersive user experiences will become more commonplace and assessing these experiences will gain importance. This will be an upcoming challenge for the user experience (UX) and quality of experience (QoE) community.

While affordable consumer VR systems are a current trend, the question of how to assess experience in immersive VE is a continuing debate in the HCI community. Bowman et al. [2] presented an overview on VE usability

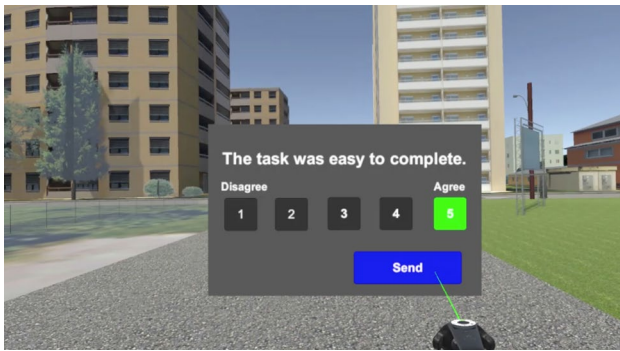


Fig. 1 Questionnaire embedded in the virtual environment

evaluation methods and Livatino and Koeffel [21] proposed guidelines for evaluating VR applications. Regarding measurements to evaluate UX in VE both subjective methods, e.g. questionnaires (cf. [7, 20, 35]), as well as objective methods, e.g. eye-tracking (cf. [32]) or biophysical measurements (cf. [18]) have been proposed. For subjective self-reported data, current UX evaluations in immersive VE often use post-test questionnaires answered after the VR experience in the real world. But especially in study designs where participants have to answer multiple questionnaires, leaving the VR experience to answer the questions in the real world is time consuming and might influence participants' sense of presence, immersion or involvement, and thus UX. Moreover, experience measurements applied after the VR experience (e.g. post-test questionnaires or interviews) rely on memory recall and thus are not desirable for separate evaluations of multiple subsequent tasks.

Although methods for evaluating experience in VE have been proposed, the challenge of how to integrate questionnaires into an immersive VE has not been systematically researched so far. In earlier work we proposed VRate (Fig. 1), a virtual questionnaire asset that can be embedded directly in a VE [24]. In this work we extended the VRate asset and investigated how such virtual questionnaires can be applied to UX evaluations in VR and whether ratings in VE are comparable to ratings in paper-pencil questionnaires.

Therefore, in this work we address the following research questions:

- RQ1: Do questionnaires in an immersive virtual environment elicit similar ratings to ratings gathered in paper-pencil questionnaires?
- RQ2: Are questionnaires in an immersive virtual environment perceived as practical by users? How do users evaluate their experience of answering questionnaires within an immersive virtual environment?
- RQ3: What is the best mounting position of a questionnaire in an immersive virtual environment?

- RQ4: Should questionnaires be embedded in the same or in a separate virtual environment? How does such placement influence the user experience and especially users' feeling of presence?

This paper is structured as follows. Firstly, we discuss related work regarding experience factors and evaluation methods for immersive VE. Secondly, we describe two studies conducted to find out if ratings (e.g. experienced presence) collected in VE are comparable to ratings obtained in paper-pencil questionnaires and how to integrate questionnaires into an immersive VE. We then discuss the questionnaire design as well as effects of placement (in-scene environment vs. dedicated environment) and mounting position, and provide suggestions for future implementations of questionnaires in immersive VE.

Related work

Evaluating user experience in fully immersive environments is different to evaluating desktop graphical user interfaces and thus poses special challenges and constructs that need to be taken into account [2].

Presence, the “feeling of being there”, was proposed as an important aspect of user experience in VE which has led to the development of questionnaires to measure presence (e.g. [20, 37]).

On the contrary, the term immersion has mostly been defined twofold in the literature: Either as denoting the objective properties of the VR system that influence the sense of presence [30, 31] or as “a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences” [37].

Although presence is seen as a key factor, several authors argue that focusing on presence does not reflect the holistic UX in immersive VE. Chertoff et al. [7] proposed the “Virtual Experience Test”, a survey instrument to measure holistic VR experiences based on a sensory, cognitive, affective, active, and relational dimension. Tcha-Tokey et al. [35] focused on concepts such as engagement, immersion, flow, skill, emotion, usability, technology adoption, judgement and experience consequence. Takatalo [33] proposed the “EVE Experience Questionnaire” that consist of 19 scales for measuring experience in VE. These include 11 scales for physical presence, interaction and flow and 8 scales for different feelings.

As a broad range of measurements for UX in VE and presence have been proposed, the question arises how to integrate such measurements in evaluations of VE. Currently, scientific literature reports on different approaches for evaluating experience in VE. For example: asking questions

orally while the participant is still wearing a HMD [23], post-experience questionnaires answered outside the VE combined with biophysiological measurements [18], eye tracking [32], etc.

An early work on usability evaluations in virtual environments was performed by Bowman et al. [2]. Apart from its detailed overview on possible evaluation methods, an important issue is raised in this work: a facilitator interacting with the participant, e.g. by talking to the participant, can influence presence. Thus, ideally there should be no interaction between the participant and the facilitator during the tasks. If presence needs to be measured, then asking questions verbally during the experience might not be suitable.

Another important issue, for all kinds of user evaluations and not specific to VR experiences, is that post-experience assessments rely on memory recall and thus might be inaccurate or incomplete [4, 15, 32]. One solution could be to use only objective measures within the VR environment (i.e. biophysical signals, eye tracking, task completion time etc.). However, as argued by Tcha-Tokey et al. [35], subjective measurements are used to understand the user's point of view and thus applying only objective measurements is not sufficient for capturing the holistic UX.

Furthermore, it is crucial for scientific experiments to consider the influence and effects of the design of measurements on the collected data. In related work, the influence of web-based questionnaires versus paper-pencil questionnaires has been studied. On the one hand, data reported by De Beuckelaer and Lievens [10] provides strong indication for measurement equivalence between internet and paper-pencil surveys, and Davidov and Depner [9] concluded based on a study comparing online and paper-pencil questionnaires that "the methods of measurement are essentially invariant" [9, p. 20]. On the other hand, Kelly et al. [19] found significant differences between electronic questionnaires and paper-pencil questionnaires as well as the interview mode. In the domain of subjective workload assessment different methods were compared using a variety of indicators [16]. This study focused on usability aspects of the methods e.g. the acceptance of the methods by operators.

A main contribution of this paper is insights into the design space of integrating questionnaires in VE. Especially the mounting position of user interface elements in VE needs consideration. Fagerholt and Lorentzon [11] discussed the design space for user interface elements in games. The authors defined two dimensions: fictional versus non-fictional and spatial versus non-spatial. In the present work, we follow the dimensions proposed by Fagerholt and Lorentzon [11] when investigating different mounting positions.

In the Quality of Experience (QoE) community, evaluation designs that require multiple ratings of a large number of samples are used. Therefore, approaches for using in-situ ratings have been proposed. For example, Buchinger et al. [5] investigated the usage of data gloves for rating time-continuous media in subjective multimedia assessment. In the subjective assessments of immersive 360-degree images, Upenik et al. [36] used a "3D immersive voting menu" embedded in the virtual environment. Although the voting menu has a similar aim to the VR questionnaire, the design and its possible effects on data gathered through the voting menu are not reflected in their work.

Of high relevance is the recent work of Schwind et al. [29], who investigated the influence of presence in VR questionnaires and compared it against a similar electronic questionnaire administered outside of the VE. The authors concluded that answering questionnaires inside VR does not influence presence and can increase consistency of questionnaire data.

In Regal et al. [24], we described the first technical implementation of a questionnaire asset for Unity. Building on this work we want to thoroughly investigate the usage of such a questionnaire in VR and the effects on ratings and user experience. To the best of our knowledge, a comparison of an immersive questionnaire embedded in VR against a paper-pencil questionnaire and the influence of both modes on presence measurements have not been analyzed so far. Further, the UX, suitability and optimal design of an immersive questionnaire embedded in VR have not been investigated in related work so far. Therefore, this work provides first insights into the design space of an immersive questionnaire embedded in VR.

Setup and methodology

In order to answer the research questions, we conducted two separate studies: one study comparing user experience ratings assessed by means of traditional paper-pencil questionnaires with a rating interface embedded in the VE, and a second study for comparing different VR questionnaire interfaces and VR questionnaire environments. Even though the study designs differ, the studies shared a variety of components such as the developed VE, the tasks the participants had to fulfil, and the technical setup. As each study took on average 80 min, combining both studies would have led to strong fatigue of the participants and with that to unreliable results. The components that both

Table 1 Experimental tasks

Task ID	Name	Description
Task 1	Move benches	Navigate along a path marked with colored dots, disable the raycast beam and then grab benches and move them to different marked spots
Task 2	Plant trees	Navigate along a path marked with colored dots and plant trees on defined spots by selecting a tree in a menu and planting the tree by pressing the trigger button. Increase/decrease the avatar size to also plant trees with different avatar sizes
Task 3	Take pictures	Navigate along a path marked with colored dots and take pictures at defined spots by pressing the wheel button on the controller
Task 4	Move cars	Navigate along a path marked with colored dots, disable the raycast beam and then grab cars and move them to different marked spots

studies have in common will be explained in the following section.

Virtual environment

The VE used in both studies is a realistic 3D model of a university campus and the surrounding environment. For implementation of the immersive VE, the game engine Unity¹ was used, which allows interaction with the virtual environment and easy integration of fully immersive head-mounted displays (HMD).

To ensure that participants have a similar experience with the system, we implemented a series of standard tasks (see Table 1 for a description of each task). The tasks were defined to cover the most common 3D interface interaction techniques: selection and manipulation, travel and system control (cf. [3]).

VR rating interface

Although different modalities for implementing a questionnaire (voice based, 3D interfaces, etc.) are possible, we decided to use a 2D graphical user interface (GUI) for the following reasons. On the one hand, 2D GUIs are familiar to users from interaction with standard desktop programs and this reduces complexity for novice VR users. Also, 2D GUIs are comparable with electronic questionnaires and from a technical point of view 2D GUIs can be easily integrated with existing survey frameworks, such as limesurvey.² Another important reason is that they can be compared to paper-pencil questionnaires.

Therefore, in this work we used a questionnaire asset for VE which we developed, the VRate asset [24]. The VRate asset (available as an open source project here: <http://vrate.tech-experience.at/>) provides a questionnaire interface for

¹ <https://unity3d.com>.

² <https://www.limesurvey.org>.



Fig. 2 M1 hand-mounted: P1 in-scene



Fig. 3 M2 head-up display: P1 in-scene

integrating a questionnaire into an immersive VE with Unity. The VRate asset consists of a user interface canvas that can be manipulated with a graphic raycast beam attached to the right hand controller [17]. We proposed that users operate the beam for teleportation and for questionnaire interaction with the right hand, yet participants were allowed to switch controllers if another mapping was preferred. For triggering actions (e.g. button press) the controller's physical trigger button was used. For details we refer to Regal et al. [24]. Pressing a button provided acoustic feedback through loudspeakers and vibrational feedback from the controllers.

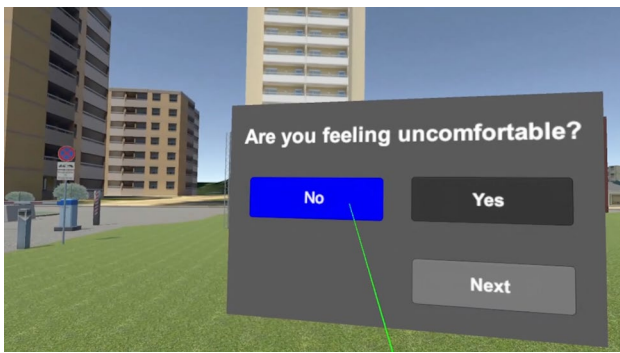


Fig. 4 M3 billboard: P1 in-scene

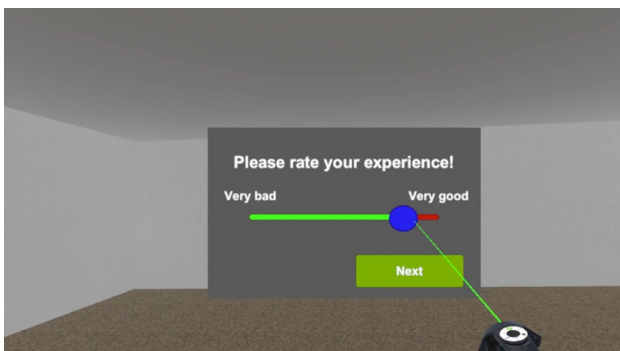


Fig. 5 Separate virtual questionnaire environment

We extended the VRate asset that provides a billboard-like rating canvas with further mounting positions for the VR questionnaire to investigate the best positioning of a questionnaire in a VE (RQ3), following the dimensions proposed by Fagerholt and Lorentzon [11]. The following mounting positions were added: M1) Hand-Mounted (Fig. 2): questionnaire attached to the controller; M2) Head-Up Display (HUD) (Fig. 3): questionnaire attached to the users head camera and thus always in sight; and M3) Billboard (Fig. 4): questionnaire displayed in a static position in the VE not affected by the user's movement. Each position has functional advantages and disadvantages (e.g. readability, visibility, focus of attention). For example M2 HUD is easily visible but blocks the field of view, M3 Billboard requires effort from the user to position themselves in front of it, M1 Hand-Mounted might be ignored by the user, etc. Thus, we assume that each position has a different effect on user experience, which we aim to investigate in this work.

To investigate whether questionnaires mounted to position M1, M2 or M3 should be embedded in the same or in a separate virtual environment (RQ4), we defined two placements of the VR questionnaire:

- P1: In-Scene, where the questionnaire is displayed in the same VE where users completed their tasks (e.g. Fig. 2)

- P2: in a dedicated virtual room where users are teleported after each task (Fig. 5).

Technical setup

We used a commercial off-the-shelf VR system, the HTC Vive (<https://www.vive.com>). To control the avatar in the VE, we used the HTC Vive controllers. We decided to use the HTC Vive, as its controllers offer an intuitive way of interacting with our system and the tracking, especially the possibility for walking around in VR, called room-scaling, is very promising. The study was conducted at two locations. In the tx.Lab of the AIT Austrian Institute of Technology GmbH, the tracking area was 3×4 m. The HTC Vive was used with an Alienware gaming PC with the following specifications: Intel Core i7-5820K 3.8 GHz, 16 GB RAM, NVIDIA GeForce GTX 980 Ti with 6 GB DDR5. In the Quality and Usability Lab of the Technical University of Berlin, the tracking area was 2.5×4 m. The HTC Vive was used with a gaming PC equipped with an Intel Core i7-7700K 4.2 GHz, NVIDIA GeForce GTX 1070 graphics card, and 16 GB of RAM.

Adjustment and training

At the beginning of each experiment, the HTC Vive HMD was adjusted to fit the participant comfortably. This action was followed by an introduction to the HTC Vive and a training session. Participants received a verbal explanation by the study facilitator on the different interaction possibilities the virtual world provides. This included an explanation of buttons, which were as follows: (1) trigger: index finger used to execute actions such as teleportation or moving objects; (2) select: big round selection button at the thumb to select trees. Additionally an icon was shown on the interface to change the avatar perspective (increase/decrease size of avatar) or to select a photo camera (see Table 1 for description of Task 2 and 3); (3) menu: small menu button (above the touchpad) to switch between interaction modes (tree, camera, nothing). After the explanation, participants could try out the controls during a training session. In this session they learned how to move in the virtual world; how to move objects; to plant trees; to change the height of the camera; to take a photo; and finally how to answer the questionnaire within the VE.

Within the training scenario the following interaction options were possible: locomotion by movement (walking within the tracking area); locomotion by teleportation with a raycast-beam; grabbing and moving objects; planting trees; taking a picture; enlarging and reducing avatar size; and interacting with a UI canvas. Participants could perform

each interaction as often as they wanted during training. If the participant felt confident enough, the training was ended and the participant had to take a 5 min break.

Comparison of the VR rating interface with paper–pencil questionnaires

This experiment investigated whether questionnaires administered in virtual environments and paper–pencil questionnaires administered in the real world environment elicit different responses.

Study design

We used the IGroup presence questionnaire (IPQ), which was designed to measure sense of presence in virtual environments [25]. The sense of presence, i.e. “the subjective experience of being in one place or environment, even when one is physically situated in another” [27], was considered to be an important aspect of the UX as it is influenced significantly by the visual representation of the virtual environment. The IPQ, initially developed in German, has been validated in two large studies with approximately 500 participants.

The questionnaire has three subscales and one global presence rating. The subscales are as follows [28]:

- *Spatial Presence* the sense of being physically present in the VE
- *Involvement* measuring the attention devoted to the VE and the involvement experienced
- *Experienced Realism* measuring the subjective experience of realism in the VE

After training, the participants first performed the 4 tasks indicated in Table 1 and rated their experience using either the VR Questionnaire (implemented with the VRate asset) in the VE or a paper–pencil version in the real world for all four conditions. Then, the participants performed the tasks again using the other rating method. The order in which the tasks and the rating methods were presented was counterbalanced to account for any possible learning and order effects.

Participants

The user study was conducted in the labs of TU Berlin. A total of 27 people participated (13 participants were women). The average age was 29.92 years (SD 6.86 years, min. 18 years, max. 47 years). Participants were recruited from the participant database of TU Berlin and received remuneration for taking part in the study.

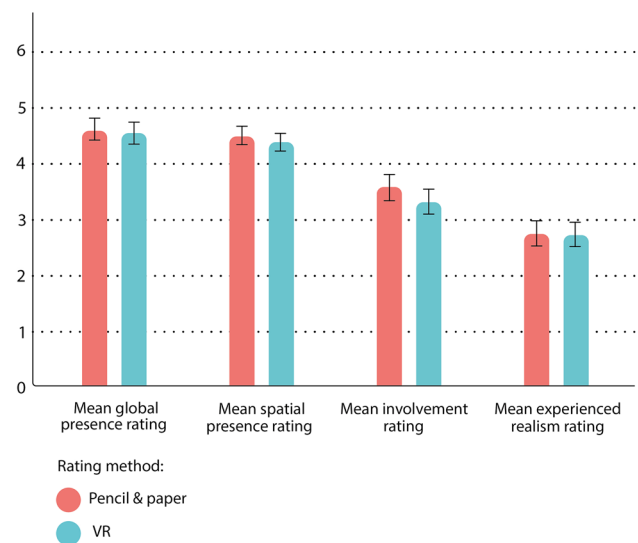


Fig. 6 Mean values of presence, spatial presence, involvement and experienced realism for all participants and tasks (benches, trees, pictures, cars). Error bars denote the 95% confidence interval

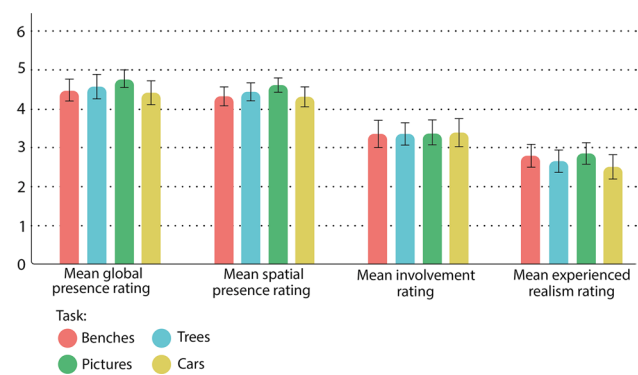


Fig. 7 Mean values of presence, spatial presence, involvement and experienced realism for all participants and rating method (paper-pencil and in VR). Error bars denote the 95% confidence interval

Twelve participants had no prior experience with VR, five described themselves as VR beginners, eight as advanced VR users and one as VR expert.

Results

A repeated measures ANOVA was conducted with the independent variables task (cf. Table 1—benches, trees, pictures, and cars) and rating method (paper–pencil and in VR) and the dependent variables presence, involvement, experienced realism, and spatial presence (IPQ dimensions)—see Figs. 6 and 7.

For the global presence, the ANOVA did not yield a significant difference for the rating method ($F_{(1,26)} = .20$, $p = .65$, $\eta_G^2 = .01$). The mean values for global presence are

very similar for both rating methods (Paper–pencil: $M = 4.59$, $SE .16$; VR: $M = 4.54$, $SE .14$).

Additionally, there were no significant differences for global presence and the different tasks ($F_{(3,78)} = 1.70$, $p = .17$, $\eta_G^2 = .06$). The average values for global presence are similar for all tasks (benches: $M = 4.50$, $SE .17$; trees: $M = 4.57$, $SE .20$; pictures: $M = 4.78$, $SE .13$; cars: $M = 4.43$, $SE .19$). For the spatial presence there were no significant differences for the rating method ($F_{(1,26)} = 2.84$, $p = .10$, $\eta_G^2 = .10$). The mean values for spatial presence are similar for both rating conditions (Paper–pencil : $M = 4.49$, $SE .14$; VR: $M = 4.37$, $SE .13$). Furthermore, there were no significant differences for spatial presence and the different tasks ($F_{(3,78)} = 2.60$, $p = .06$, $\eta_G^2 = .09$). The average values for spatial presence are similar for all tasks (benches: $M = 4.35$, $SE .15$; trees: $M = 4.46$, $SE .13$; pictures: $M = 4.60$, $SE .14$; cars: $M = 4.32$, $SE .17$). For involvement a significant difference for the rating method ($F_{(1,26)} = 13.53$, $p = .001$, $\eta_G^2 = .342$) was found. The mean involvement was higher for the condition pencil and paper (Paper–pencil : $M = 3.52$, $SE .21$; VR: $M = 3.28$, $SE .20$).

Additionally, there were no significant differences for involvement and the different tasks ($F_{(3,78)} = .06$, $p = .98$, $\eta_G^2 = .01$). The mean values for involvement are similar for all tasks (benches: $M = 3.40$, $SE .21$; trees: $M = 3.37$, $SE .20$; pictures: $M = 3.42$, $SE .20$; cars: $M = 3.42$, $SE .25$). For experienced realism, no significant differences for the rating method ($F_{(1,26)} = .07$, $p = .79$, $\eta_G^2 = .01$) was found. Additionally, there was no significant differences for experienced realism and the different tasks ($F_{(3,78)} = 2.36$, $p = .08$, $\eta_G^2 = .08$). The mean values for experienced realism are similar for all tasks (benches: $M = 2.82$, $SE .19$; trees: $M = 2.67$, $SE .18$; pictures: $M = 2.86$, $SE .19$; cars: $M = 2.51$, $SE .22$).

VR rating interface comparison

In the first study we could successfully show that questionnaire answers in VR lead to similar ratings as questionnaires answered using paper–pencil (RQ1). Therefore, we conducted a second study to answer the remaining research questions.

Study design

In the second study, we followed a mixed study design. We investigated the VR questionnaire mounting position (M1, M2, M3) as within-subject factor and placement environment (P1, P2) as between-subject factor.

Our experiment followed the hereby described procedure. First participants had to read information and sign a letter of consent. Afterwards, participants were given a short explanation of the VR hardware (tracking area, controllers,

and head mounted display). Subsequently, participants performed the training tasks described in “[Adjustment and training](#)” section.

After the training was completed, the experimental part of the study started. Participants were asked to perform tasks (Table 1) in the VE and rate their experience afterwards using the VR Questionnaire (implemented with the VRate asset). Due to the similarity of task 1 and 4, in this study only the task 1, 2, and 3 were performed by the participants. The number of conditions was reduced to avoid fatigue of the participants. Consequently, the duration of the second study was similar to the one of the first study. After a participant finished the task, the written information “Task Completed” was displayed in the VE. The message was triggered by the facilitator if the task was completed. Afterwards, the participant was allowed to freely explore the environment for the remaining time. This procedure was chosen to ensure that every participant spent roughly the same time in the VE. Five minutes after the task started, the VR questionnaire was displayed to the participants inside the immersive environment.

After each task, the VR questionnaire was presented to the participant in a different mounting position (M1, M2, M3). The order of mounting positions was counterbalanced. The VR questionnaire was placed either in the VE (P1) or in a dedicated rating room (P2). The placement was kept constant for each participant and varied between participants. Additionally, we logged the time participants needed to fill in the VR questionnaire.

When all tasks were completed, participants were instructed to take off the HMD and to answer two post-experience questionnaires displayed on a tablet device in the real world. To answer RQ2, we used the UMUX [12] questionnaire to measure the user experience when answering the VR questionnaire. For RQ4, we used the ITC SOPI [20] questionnaire to measure immersion and presence.

Subsequently, participants were asked qualitative questions on how they perceived the questionnaire. The following questions were asked : (1) “Did you like the implementation of the questionnaire in the virtual environment?”, (2) “Did you encounter any problems using the questionnaire in the virtual environment?” and (3) “Did you experience problems with the legibility of the questionnaire in the virtual environment?” Furthermore, participants were asked to rate each mounting position (M1, M2 and M3) on a 7-point Likert scale from 1 very bad to 7 very good: (4) “How did you like the following methods: Hand, anchored on the controller, Head Up Display, anchored in the field of view and Billboard, static position in the environment”. Additionally, participants were asked to order the mounting positions according to preference: (5) “During the course of the test, you have experienced three different questionnaire variants. Please rank the variants in descending order according to

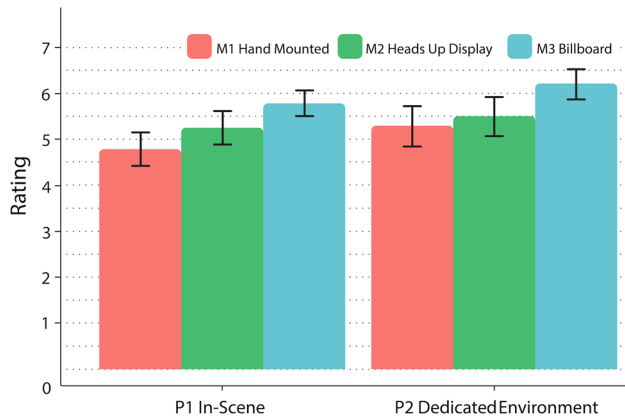


Fig. 8 Mean rating of how much participants liked the mounting positions (scale 1: very bad to 7: very good) of the questionnaires using the different implementations. Error bars show 95% within-subject confidence intervals (cf. [8])

your personal preference (place 1 = your favorite)”. Finally, we asked how the questionnaire affected their VR experience and if they would have preferred an alternative placement (P1 or P2) or a paper–pencil questionnaire outside the VE.

For qualitative data analysis, we used deductive category assignment following the procedure suggested by Mayring [22]. The answers in the interview were categorized into themes and subsequently counted how often such themes occurred. Due to the number of participants ($n = 48$) we will not provide participant identifiers for the themes, but report only the absolute count.

Participants

The user study was conducted at the AIT and QU Labs. A total of 48 people, 24 people in each lab, participated. 25 participants were women. The average age was 28.12 years (SD 5.96 years, min. 18 years, max. 52 years). Participants were recruited via Facebook and from institutional participant databases. Participants received remuneration for taking part in the study.

Regarding computer experience, most participants described themselves as advanced users (34 participants) or experts (10 participants), with only 4 participants describing themselves as beginners. Nine people had no prior experience with VR, 32 described themselves as VR beginners, 6 as advanced VR users and one as VR expert.

Results

Rating of questionnaires

Participants were asked to rate each mounting position on a 7-point rating scale (Fig. 8), which served as the independent

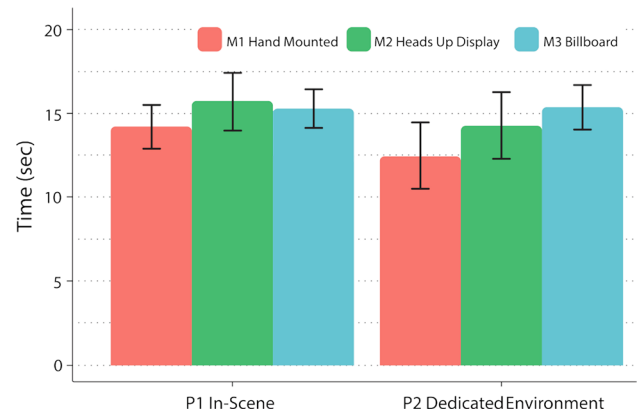


Fig. 9 Mean task completion times for filling in the questionnaires. Error bars show 95% within-subject confidence intervals (cf. [8])

variable (RATING). Two-way ANOVA with mounting position (M1 Hand, M2 HUD, M3 Billboard) as within-subjects factor and environment (P1 In-Scene, P2 Dedicated Environment) as between-subjects factor was used to analyze the data.

For the analysis, a repeated measure ANOVA was conducted. Mauchly’s test confirmed that the assumption of sphericity was met ($p = .15$). The ANOVA shows a significant main effect for mounting position ($F_{(2,92)} = 4.82$, $p = .01$, $\eta_G^2 = .07$) and no significant results for the second factor environment ($F_{(1,46)} = 2.65$, $p = .11$, $\eta_G^2 = .02$) and the interaction effect ($F_{(2,92)} = .08$, $p = .92$, $\eta_G^2 = .01$). Post-hoc analysis with paired-samples t-test and Bonferroni-corrected p values shows that M3 Billboard was rated significantly better than M1 Hand ($p_{\text{adj}} = .01$), with no significant differences between the two other possible combinations M3 Billboard—M2 HUD ($p_{\text{adj}} = .15$) and M2 HUD—M1 Hand ($p_{\text{adj}} = 1$).

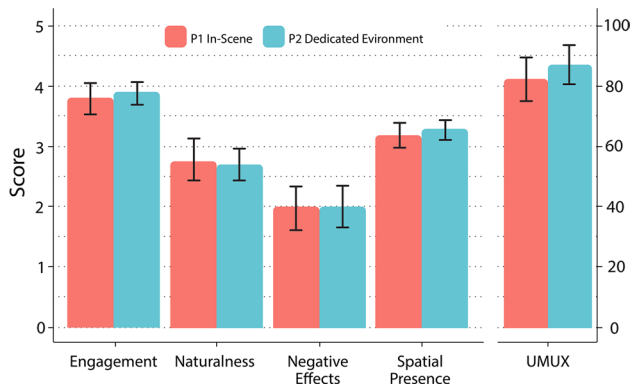
Task completion time

The overall time needed by the participants to complete the post task questionnaires (each consisting of three questions) embedded in the virtual environment in the different test conditions was logged to compare task completion times with the different mounting positions (M1, M2, M3). Figure 9 shows an overview of the required time in the different test conditions.

The ANOVA yielded no significant differences, neither for mounting position ($F_{(2,90)} = 1.12$, $p = .33$, $\eta_G^2 = .01$) nor for the environment ($F_{(1,45)} = .32$, $p = .57$, $\eta_G^2 = .01$). Due to technical problems with the logging software data from one participant is not available. Mean task completion time was 14.16 s (M1 Hand), 15.70 s (M2 HUD), 15.29 s (M3 Billboard) in the in-scene condition and 12.47 s (M1 Hand), 14.26 s (M2 HUD), 15.34 s (M3 Billboard) in the dedicated

Table 2 Number of mounting position and placement preferred by the participants

	BB	HUD	Hand	Scene	Room
Scene	11	4	9	21	3
Room	11	7	6	16	8
Total	22	11	15	37	11

**Fig. 10** Scores on the different ITC-SOPI scales (1...5) and UMUX (0...100). Error bars show 95% confidence intervals (between-subjects)

environment. It is noteworthy, when analysing the logged user behaviour, that the mounting position which was rated worst (M1 Hand) had the shortest completion time.

Preference for methods

When asking which method participants would prefer, a similar pattern emerged with 22 participants preferring M3 Billboard, 15 participants M2 HUD, and only 11 participants preferred M1 Hand. An interesting trend is shown when cross-tabling the preferred environment (Table 2). Whereas almost all participants who used the questionnaires in the scene also stated that they would prefer this method, the pattern for the participants that used the dedicated questionnaire environment is different: Here only two-thirds of participants would prefer the in-scene method. We think these results might be explained by a familiarity bias.

Presence and user experience

Presence was measured using the ITC Sense of Presence Inventory (ITC-SOPI), which distinguishes between four different dimensions of presence. Figure 10 shows that the scores are almost identical for the two environments, and statistical analysis using ANOVA did not find significant differences for any of the presence dimensions (Engagement: $F_{(1,46)} = 0.39$, $p = .54$; Naturalness: $F_{(1,46)} = 0.16$,

$p = .69$; Negative Effects: $F_{(1,46)} = 0.01$, $p = .95$; Spatial Presence: $F_{(1,46)} = 0.42$, $p = .52$). This is an indication that the two questionnaire environments influenced the presence in a similar way. Also, the results for user experience (as measured by the UMUX) are very similar. As can be seen in Fig. 10 the means of the different scales are almost identical, and 95% confidence intervals do strongly overlap. ANOVA does not show any significant effect of the environment on the user experience measured by the UMUX ($F_{(1,46)} = 0.72$, $p = .40$).

Interview

When looking at the ranking order, M3 Billboard was most often ranked first place followed by M2 HUD and M1 Hand, which was most often ranked third place. Thus, overall we can see a tendency in preference for M3 Billboard over M2 HUD. In the interview different explanations for this preference were provided.

For M3 Billboard, 11/48 participants mentioned that it felt most realistic, due to its static position in the VE. To quote one participant: “The Billboard was placed realistically in the space, you could position yourself optimally and see the surroundings well.” In contrast 5/48 participants stated that a static billboard is not flexible as it stays in the same position and can be “lost” when moving away. For example, one participant mentioned that “the Billboard could be lost, which is a bit impractical in some scenarios.” And a further participant mentioned: “A disadvantage is that you have to search for the billboard first”.

Another advantage mentioned by the participants is that the billboard allows for looking away and is not attached to the body, thus it was perceived as less annoying (10/48). As stated by one participant: “With the billboard it is my “decision” to look at it and work with it.” Also, 9/48 participants liked that the billboard allows for active positioning to optimize readability and usability. In contrast, 5/48 participants stated that active positioning is cumbersome. 4/48 participants stated that the billboard is boring.

For M2 HUD, 17/48 participants stated they preferred that the questionnaire is placed directly in the field of view and thus cannot be overlooked and does not need active positioning, as for example mentioned by one participant: “HUD is in 1st place because it was obviously visible and the positioning was automatic”. In contrast, 15/48 participants stated that placement of the VR questionnaire in the field of view is annoying. For example, a participant stated: “I felt less free and it’s like a board in front of your head.” One participant mentioned “With the HUD you couldn’t look around anymore and it was therefore unnatural.” 3/48 participants stated that the HUD is little effort. A disadvantage mentioned by 3/48 participants is that the HUD moves during answering due to involuntary head movement.

Similarly, for M1 Hand 11/48 participants mentioned problems with raycast due to involuntary movement of the hand. Also, 8/48 participants criticized that two hands are needed for interaction. 3/48 participants stated that M1 Hand is most intuitive and flexible and 4/48 participants mentioned positively that the questionnaire is attached to the avatar, but is less obtrusive than the HUD. 4/48 participants reported that they had difficulty noticing when the hand questionnaire was displayed. For example, one participant stated: “The hand questionnaire would be better if it became better noticeable. Since I didn’t notice the hand questionnaire I ranked it in 3rd place.”

Regarding placement of the questionnaire (in-scene environment vs. dedicated environment) no significant difference in the scales of ITC SOPI was found, but most participants (38/48) stated that they would prefer answering the questionnaire inside the same VE. 17/48 participants stated that switching the environment would break immersion and 12/48 participants mentioned a negative effect due to the (sudden) teleportation and the need for orientation in the new questionnaire environment. For example, one participant stated that “[...] it would irritate me to be automatically teleported without warning. The constant getting out of the scene will also affect the experience”.

8/48 participants stated that in a neutral environment no new experiences are likely to influence the answers, and 2/48 stated explicitly that inside the same VE more distraction is likely to occur. “I prefer the room because there’s less distraction.” one participant said. One participant mentioned that “no new impressions overwrite the old ones”.

Discussion

In this section we discuss the gathered insights from our comparison of virtual and paper–pencil questionnaires and how to design questionnaires embedded in VE. Subsequently, we discuss limitations of our research and draw implications for future work.

RQ1: Comparison VR questionnaires versus paper–pencil

To compare the ratings performed in VR and paper–pencil questionnaires, the scales of the IPQ were compared within participants for both rating methods. We found no significant differences for global presence (rating method and tasks). Similar results were achieved for spatial presence and experienced realism. For involvement there was a significant difference between the two different rating methods. Involvement ratings were higher for ratings performed with paper and pencil in the real world. We assume that this could be due to the fact that the questions on the IPQ involvement

scale aim to measure how e.g. aware/attentive participants are towards the real world (“I was not aware of my real environment.”). By asking questions like this while participants were still in VR, the attention of participants was directed towards the real world and therefore also resulted in lower average involvement ratings. This also correlates with recent work from Schwind et al. [29], who also found no influence for measured presence for in ratings administered within and outside VR environments.

In general, it seems that the ratings gathered in VR are comparable to the rating gathered via paper and pencil. Questions and rating scales that directly divert the attention away from the VR experience have to be interpreted carefully when gathered in VR.

To verify the results of the first study we included the question “Would you have preferred to answer the questionnaire outside the immersive VE on paper” in the interview in the second study. In the interview we found clear evidence that answering the questionnaire inside the VE is preferred as only 2 of 48 participants would have preferred to answer the questionnaire outside the VE on paper. As an explanation participants mentioned mostly comfort (fitting and removing the HMD is uncomfortable and cumbersome) and that switching to reality for answering questions would disrupt immersion and experience. The two participants that would have preferred a questionnaire outside the VR mentioned that a break from VR would have lessened strain especially for lengthy VR evaluations.

RQ2: UX of the VR questionnaire

Overall the VR questionnaire was received well by the participants. The results of the UMUX questionnaire for both conditions (inside the scene and outside the scene) were very positive (> 80 cf. Fig. 10) thus this suggests that the VR questionnaire was perceived as easy to use.

Also, in the qualitative interview hardly any problems regarding usability when interacting with the VR questionnaire were reported. Mostly, the implementation of the slider rating mechanism was criticized by participants (6/48). Readability was also perceived well by the participants, only 3/48 participants (slightly) agreed with the question “Did you have problems with readability of the questionnaire in VR”.

Thus, regarding RQ2 (Are questionnaires in an immersive virtual environment perceived as practical by users?) we can summarize that, overall, answering questionnaires in VR was a positive experience for the participants. This is especially important as a negative experience when answering the VR questionnaire, for example due to interface design issues, can negatively influence the overall VE experience and thereby the user scores gathered through it.

RQ3: Mounting position of the VR questionnaire

Regarding RQ3 “What is the best mounting positioning of a questionnaire in an immersive virtual environment?” (M1 hand-mounted, M2 Head Up Display or M3 Billboard), we found that M3 Billboard was rated significantly better than M1 Hand.

A strong benefit of M3 Billboard was that it is perceived as real and natural and thus integrates well into the VE. The possibility to look away and ignore the billboard was seen as a benefit. In contrast, the billboard can also be overlooked or lost. We suggest that in evaluations where realism is most important M3 Billboard should be used, although evaluators need to be aware that billboards can be more easily overlooked or ignored than the other mounting positions. Although, M3 Billboard was rated best, all mounting positions are rated positively in the interview (score 5 to 7 in the 7-point rating scale). Thus we also think the other options (M1 Hand and M2 HUD) are suitable to use if they fit the desired experimental setup better than M3 Billboard.

M2 HUD was often criticized as too annoying as it is stuck in the field of view and does not allow looking away. On the other hand this can also be considered as a benefit as it is highly visible and cannot be lost (in comparison to the billboard). If the user is required to answer a questionnaire immediately, it is preferable to position the questionnaire as a HUD, as it must be answered to continue doing other tasks in a practical manner. Possible issues with readability (suggested by VR design guidelines—although not found in this study) can be minimized if the distance to the HUD can be actively chosen by the user or is personalized beforehand.

M1 Hand cannot be lost and does not block the field of view. This is a clear benefit that makes this mounting position especially useful if the questionnaire does not need to be answered immediately and it can also be carried around by the user. Also, M1 Hand allows the user to explore their surrounding if this is important for answering the questionnaire. On a negative side, participants stated that it is cumbersome to use both hands, and some participants did not notice that the questionnaire was displayed as the hand was outside their field of view. We suggest that vibration or acoustic feedback is provided when the questionnaire is displayed.

For the mounting positions M1 hand 11/48 participants reported difficulty when aiming with the raycast, as the questionnaire moved due to involuntarily movement of the hand when trying to perform an action. Interestingly, some participants did not notice (initially or at all) that the questionnaire can be moved by moving the hand, which led to uncomfortable postures when filling out the questionnaire. Difficulty aiming with the raycast was also reported for M2 HUD but not as often and not as severe. As a possible solution answering the VR questionnaire could also be implemented in another way, for example non-raycast

based interaction, such as direct touch or using controller buttons. For comparability we intentionally used raycast for all 3 mounting options in this study. M1 Hand could benefit particularly from other interaction methods, an aspect which needs to be investigated in future work.

RQ4: Placement of the VR questionnaire

Neither an effect on the ITC SOPI scales (spatial presence, engagement, ecological validity/ naturalness and negative effects) nor a difference in interruption of experience in the interview was found. This is a notable finding as this is in contrast to our initial assumptions and also in contrast to the preferences reported in the interview.

In the interview only 10 of 48 participants stated that they prefer a separate questionnaire environment. With a dedicated environment sudden teleportation is an issue, and in-scene questionnaires are perceived as less disruptive. A benefit of the dedicated questionnaire environment is that there are fewer distractions, and the user can focus completely on the questionnaire.

Therefore, we suggest that embedding the VR questionnaire in the same environment should be preferred, however a dedicated questionnaire environment could also be chosen if it is more suitable for the desired evaluation design. In terms of technical implementation, a dedicated questionnaire environment can be implemented as a separate scene in Unity and thus integration requires less effort than integrating a questionnaire asset inside a scene. Therefore, due to easier implementation, we suggest using a separate virtual questionnaire environment in study designs where multiple different scenes are evaluated, for example in QoE studies (cf. [26]) where different scenes with different renderers are evaluated.

Limitations and future work

This work is a first step towards investigating the design of VR questionnaires embedded in fully immersive VE. Although we have gathered promising results, some questions remain open for investigation in future work.

A certain limitation is the number of participants in the evaluations. We included 27 participants in the first and 48 participants in the second study. While, this is a suitable sample size to gather first insights, more evaluations are needed to confirm the results and suggestions presented in this paper. Another limitation is the VR equipment (HTC Vive) used, the virtual 3D environment used and the type and number of tasks (cf. [14]) conducted by the participants. It is possible that other 3D environments and/or other tasks could have stronger effects on the conducted measurements, especially for presence.

Also, it is uncertain whether questionnaire inside VR can result in better data quality. For example, although measurements of visual quality can be done directly in the scene without the need to remember the quality in post experiment assessment our first study shows no difference in data quality for the measured construct presence (subscales: global presence, spatial presence, and experiences realism) between paper–pencil and questionnaire inside the VE. Nevertheless, similar to De Beuckelaer and Lievens [10] large scale evaluations are needed to compare data quality from the VR questionnaire against paper–pencil or electronic questionnaires administered in the real world. Therefore we would like to emphasize that the VRate asset for Unity [24] is open source (available at <http://vrate.tech-experience.at/>) and can be used by other researchers to investigate the issues addressed in this work.

Another possible limitation is the current implementation of the VRate asset. In this work we decided to use a 2D user interface with graphic raycast interaction. Although this is a common form of interaction in VE, a broad range of other interaction modalities for 3D user interfaces exist (cf. [3]). In future work, other implementations of interaction modalities for VR questionnaires should be investigated. Immersive VE also has the potential to build playful VR questionnaires, similar to playful web surveys proposed by Harms et al. [13], Takhtamysheva and Smeddinck [34].

A further possible limitation is that a repeated measures (within-subject) design was used in the first study. Between versus within-subject design is an often debated issue in scientific literature (cf. [6]), as there are benefits for both approaches. In the first study we aimed to investigate the influence of environment on the scales itself. As we wanted to eliminate possible interpersonal effects we conducted a within-subject study with repeated measures of the same construct (presence). In the second study we aimed to investigate the influence of questionnaire placement on global presence experience, therefore we used a between-subject design, and thus evaluated twice the number of participants compared to the first study.

Summary

In this work, we presented first insights into the design of a questionnaire embedded inside a virtual environment. We extended the VRate [24] questionnaire asset and compared it against a classic pencil and paper approach. Moreover, we investigated user experience and perceived suitability of answering a VR questionnaire inside an immersive VE. We also investigated the effects of different design decisions like mounting position and placement.

We found that the ratings gathered in VR are comparable to the ratings gathered in the real world by pencil and paper

questionnaires (subscales: global presence, spatial presence, and experiences realism). Also, we found a strong indication that participants preferred answering questionnaires inside the virtual environment, mostly due to greater comfort as questionnaires can be completed without removing the head mounted display. Also, the questionnaire inside the virtual environment was perceived as easy to use. Regarding mounting position, there is a tendency that billboard is preferred although other positions (hand-mounted or head up display) might be suitable as well, depending on the desired goals. Based on the gathered insights, we have discussed pros and cons of the different placement and mounting options and provide suggestions for future implementations of questionnaires embedded in virtual environments.

Acknowledgements The authors would like to thank all project partners for their valuable work and contributions as well as for their effort. Furthermore, the authors would like to especially thank Volker Settigast (Fraunhofer Austria) for providing the 3D environment model, Arno de Meyer for helping with the data recording, Helmut Schrom-Feiertag for feedback on the questionnaire implementation and Valentin Gattol for proofreading. This work has been partially funded by the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit) in the “Mobilitaet der Zukunft” program under Grant No. 860210 (“VR-Planning”) and Grant No. 865214 (“FX Future eXperience”).

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding authors state that there is no conflict of interest.

References

1. Anthes C, García-Hernández RJ, Wiedemann M, Kranzlmüller D (2016) State of the art of virtual reality technology. In: Proceedings of the IEEE aerospace conference, pp 1–19. <https://doi.org/10.1109/AERO.2016.7500674>
2. 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. <https://doi.org/10.1162/105474602760204309>
3. Bowman DA, Kruijff E, LaViola JJ Jr, Poupyrev I (2001) An introduction to 3-D user interface design. *Presence Teleoper Virtual Environ* 10(1):96–108. <https://doi.org/10.1162/105474601750182342>
4. Bruun A, Ahm S (2015) Mind the gap! Comparing retrospective and concurrent ratings of emotion in user experience evaluation. In: Abascal J, Barbosa S, Fetter M, Gross T, Palanque P, Winckler M (eds) *Human-computer interaction – INTERACT 2015. Lecture Notes in Computer Science*, vol 9296. Springer, Cham
5. Buchinger S, Robitza W, Nezveda M, Hotop E, Hummelbrunner P, Sack MC, Hlavacs H (2014) Evaluating feedback devices for time-continuous mobile multimedia quality assessment. *Signal Process Image Commun* 29(9):921–934. <https://doi.org/10.1016/j.image.2014.07.001>
6. Charness G, Gneezy U, Kuhn MA (2012) Experimental methods: between-subject and within-subject design. *J Econ Behav Organ* 81(1):1–8. <https://doi.org/10.1016/j.jebo.2011.08.009>

7. Chertoff DB, Goldiez B, LaViola JJ (2010) Virtual experience test: a virtual environment evaluation questionnaire. In: Proceedings of the IEEE virtual reality conference, pp 103–110. <https://doi.org/10.1109/VR.2010.5444804>
8. Cousineau D (2005) Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutor Quant Methods Psychol* 1(1):42–45. <https://doi.org/10.20982/tqmp.01.1.p042>
9. Davidov E, Depner F (2011) Testing for measurement equivalence of human values across online and paper-and-pencil surveys. *Qual Quant* 45(2):375–390. <https://doi.org/10.1007/s11135-009-9297-9>
10. De Beuckelaer A, Lievens F (2009) Measurement equivalence of paper-and-pencil and Internet organisational surveys: a large scale examination in 16 countries. *Appl Psychol* 58(2):336–361. <https://doi.org/10.1111/j.1464-0597.2008.00350.x>
11. Fagerholt E, Lorentzon M (2009) Beyond the HUD. User interfaces for increased player immersion in FPS games. Master of science thesis, Chalmers University
12. Finstad K (2010) The usability metric for user experience. *Interact Comput* 22(5):323–327. <https://doi.org/10.1093/iwc/iwt005>
13. Harms J, Biegler S, Wimmer C, Kappel K, Grechenig T (2015) Gamification of online surveys: design process, case study, and evaluation. *Hum Comput Interact* 9296:219–236. https://doi.org/10.1007/978-3-319-22701-6_16
14. Hassenzahl M, Kekez R, Burmester M (2002) The importance of a software's pragmatic quality depends on usage modes. In: Proceedings of the 6th international conference on work with display units, vol 3, pp 275–276
15. Hassenzahl M, Sandweg N (2004) From mental effort to perceived usability: transforming experiences into summary assessments. In: Chi04 extended abstracts on human factors in computing systems. ACM, pp 1283–1286. <https://doi.org/10.1145/985921.986044>
16. Hill S, Iavecchia H, Byers J, Bittner A, Zaklade A, Christ R (1992) Comparison of four subjective workload rating scales. *Hum Factors* 34(4):429–439
17. Hinckley K, Pausch R, Goble JC, Kassell NF (1994) A survey of design issues in spatial input. In: Proceedings of the 7th annual ACM symposium on User interface software and technology, pp 213–222. <https://doi.org/10.1145/192426.192501>
18. Janßen D, Tummel C, Richert A, Isenhardt I (2016) Towards measuring user experience, activation and task performance in immersive virtual learning environments for students. In: Proceedings of the international conference on immersive learning, pp 45–58. https://doi.org/10.1007/978-3-319-41769-1_4
19. Kelly D, Harper DJ, Landau B (2008) Questionnaire mode effects in interactive information retrieval experiments. *Inf Process Manag* 44(1):122–141. <https://doi.org/10.1016/j.ipm.2007.02.007>
20. Lessiter J, Freeman J, Keogh E, Davidoff J (2001) A cross-media presence questionnaire: the ITC-sense of presence inventory. *Presence Teleoper Virtual Environ* 10(3):282–297. <https://doi.org/10.1162/105474601300343612>
21. Livatino S, Koeffel C (2007) Handbook for evaluation studies in virtual reality. In: Proceedings of the IEEE international conference on virtual environments, human-computer interfaces, and measurement systems, pp 1–6. <https://doi.org/10.1109/VECIM.S.2007.4373918>
22. Mayring P (2014) Qualitative content analysis: theoretical foundation, basic procedures and software solution. Klagenfurt. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173>
23. Men L, Bryan-Kinns N, Hassard AS, Ma Z (2017) The impact of transitions on user experience in virtual reality. In: Proceedings of the IEEE virtual reality conference, pp 285–286. <https://doi.org/10.1109/VR.2017.7892288>
24. Regal G, Schatz R, Schrammel J, Suette S (2018) VRRate: a Unity3D asset for integrating subjective assessment questionnaires in virtual environments. In: Proceedings of the international conference on quality of multimedia experience (QoMEX), pp 1–3. <https://doi.org/10.1109/QoMEX.2018.8463296>
25. Regenbrecht H, Schubert T (2002) Real and illusory interactions enhance presence in virtual environments. *Presence Teleoper Virtual Environ* 11(4):425–434
26. Schatz R, Regal G, Schwarz S, Suette S, Kempf M (2018) Assessing the QoE impact of 3D rendering style in the context of VR-based training. In: Proceedings of the international conference on quality of multimedia experience (QoMEX), pp 1–6
27. Schubert T (2003) The sense of presence in virtual environments: a three-component scale measuring spatial presence, involvement, and realism. *Zeitschrift für Medienpsychologie* 15(2):69–71
28. Schubert T, Friedmann F, Regenbrecht H (2001) The experience of presence: factor analytic insights. *Presence Teleoper Virtual Environ* 10(3):266–281
29. Schwind V, Knierim P, Haas N, Henze N (2019) Using presence questionnaires in virtual reality. In: Proceedings of the 2019 CHI conference on human factors in computing systems: CHI '19, pp 1–12. <https://doi.org/10.1145/3290605.3300590>
30. Slater M (1999) Measuring presence: a response to the Witmer and Singer presence questionnaire. *Presence Teleoper Virtual Environ* 8(5):560–565. <https://doi.org/10.1162/105474699566477>
31. Slater M (2003) A note on presence terminology. http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/ConsortiumPublications/ucl_cs_papers/presence-terminology.htm. Accessed 25 Oct 2019
32. 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: Proceedings of the 9th international conference on virtual, augmented and mixed reality, pp 369–380. https://doi.org/10.1007/978-3-319-57987-0_30
33. Takatalo J, Nyman G, Laaksonen L (2008) Components of human experience in virtual environments. *Comput Hum Behav* 24(1):1–15. <https://doi.org/10.1016/j.chb.2006.11.003>
34. Takhtamyshva A, Smeddick J (2012) Serious questions in playful questionnaires. In: Proceedings of the international conference on entertainment computing, pp 449–452. https://doi.org/10.1007/978-3-642-33542-6_50
35. Tcha-Tokey K, Loup-Escande E, Christmann O, Richir S (2016) A questionnaire to measure the user experience in immersive virtual environments. In: Proceedings of the 2016 virtual reality international conference, pp 1–5. <https://doi.org/10.1145/2927929.2927955>
36. Upenik E, Rerabek M, Ebrahimi T (2017) On the performance of objective metrics for omnidirectional visual content. In: Proceedings of the international conference on quality of multimedia experience (QoMEX), pp 1–6. <https://doi.org/10.1109/QoMEX.2017.7965660>
37. Witmer BG, Singer M (1998) Measuring presence in virtual environments: a presence questionnaire. *Presence* 7(3):225–240. <https://doi.org/10.1162/105474698565686>

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