Investigating Effects and User Preferences of Extra- and Intradiegetic Virtual Reality Questionnaires

Nadine Wagener
University of Bremen
Bremen, Germany
nwagener@uni-bremen.de

Mareike Stamer
Volkswagen AG
Wolfsburg, Germany
mareike.stamer@volkswagen.de

Johannes Schöning
University of Bremen
Bremen, Germany
johannes.schoening@uni-bremen.de

Johannes Tümler
Anhalt University of Applied Sciences
Köthen, Germany
johannes.tuemler@hs-anhalt.de

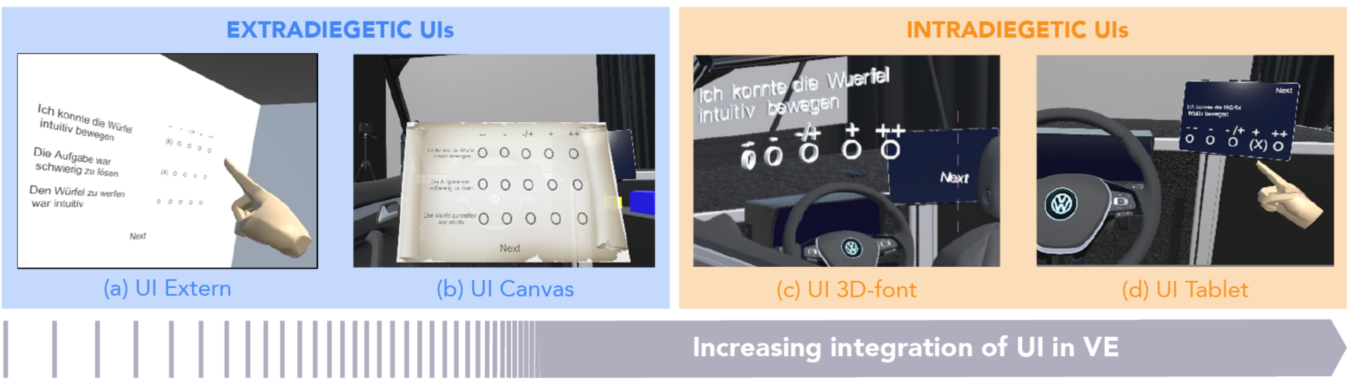

Figure 1: Two extradiegetic UIs (blue) and two intradiegetic UIs (orange) for VR-questionnaires ($_{IN}$VR-Qs). From left to right the level of integration in the VE increases. (a) *UI Extern* shows evaluation in new VE, (b) *UI Canvas* shows evaluation on paper roll attached to HMD, (c) *UI 3D-font* shows evaluation via 3D-elements, (d) *UI Tablet* shows evaluation on tablet.

ABSTRACT

Virtual realities (VR) are becoming an integral part of product development across many industries, for example to assess aesthetics and usability of new features in the automotive industry. The recording of the evaluation is typically conducted by filling out questionnaires after the study participants left the virtual environment. In this paper, we investigate how questionnaires can be best embedded within the virtual environment and compare how VR-questionnaires differ from classical post-test evaluations regarding preference, presence, and questionnaire completion time.

In the first study ($N = 11$), experts rated four design concepts of questionnaires embedded in VR, of which two were designed as *extradiegetic* and two as *intradiegetic* user interfaces. We show that intradiegetic UIs have a significantly higher perceived user experience and presence while the usability remains similar. Intradiegetic UIs are preferred by the majority.

Based on these findings, we compared intradiegetic VR-questionnaires with paper-based evaluations in a follow up study ($N = 24$). 67% of the participants preferred the evaluation in VR, even though it takes significantly longer. We found no effect on presence.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; *User interface toolkits*; *Gestural input*.

KEYWORDS

Virtual Reality (VR), User Interface (UI), Evaluation, $_{IN}$VR- questionnaires, Presence

ACM Reference Format:
Nadine Wagener, Mareike Stamer, Johannes Schöning, and Johannes Tümler. 2020. Investigating Effects and User Preferences of Extra- and Intradiegetic Virtual Reality Questionnaires. In *26th ACM Symposium on Virtual Reality Software and Technology (VRST '20), November 1–4, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3385956.3418972

VRST '20, November 1–4, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-7619-8/20/11…$15.00
https://doi.org/10.1145/3385956.3418972

1 INTRODUCTION & MOTIVATION

Virtual Reality (VR) offers opportunities to test complex products and prototypes in simple, cost-effective and efficient ways regarding design and functional usability [4]. Brade et al. [6] suggest that for testing in VR it is highly important for study participants to feel physically located in the virtual environment (VE), which means reaching a high presence [37, 44].

Research with VR experiences often rely on post-test surveys conducted after the study participants left the VE. The assessment itself, most commonly done via questionnaires [3], is thus medially and visually separated from the actual test phase in VR. Relying on these kind of post-test assessments affects the emotional state, behaviour and performance of study participants in many ways. Transitioning from VR to reality can negatively affect the performance and motivation of users [17] and can cause disorientation [15, 35]. Presence gets disrupted [39] and it takes time for participants to re-orient themselves in the new environment after a head-mounted-display (HMD) was removed and put on again [35]. Additionally, the participants need to constantly remind themselves of what they have experienced when filling out the questionnaire [15]. This causes errors in the raw data that do not accurately reflect the subject's opinion of the product or system being tested. Further studies found that spontaneous thoughts and feelings are not adequately included in usability assessments [13, 17]. Being in reality, the participants are subject to systematic biases by the experimenter or the study environment as well [30].

As a solution, embedding questionnaires directly in VR promises to mitigate the effects of classical post-test assessments [3, 30, 35]. In this paper, we refer to the integration of questionnaires in VR as $_{IN}$VR-Qs. First approaches already verify the validity of such evaluations [19, 30, 35], suggesting a reduction of the time needed for the whole study [35] and higher enjoyment [3].

Although Haas [19], Schwind et al. [35] and Putze et al [30] medially integrated the evaluation in VR, thus used $_{IN}$VR-Qs, they still visually separated the evaluation from the experienced VE. The ongoing experience got exchanged with another VE for evaluation purposes. We refer to this approach as using *extradiegetic* $_{IN}$VR-Qs [40]. However, Frommel et al. [15] stress the importance of staying close to the context of the ongoing experience when answering questionnaires. In this paper, we refer to this concept as utilising *intradiegetic* $_{IN}$VR-Qs [40]. Some approaches along this line include the thinking-aloud protocol [25] or an oral administration of questions [10], however, both can decrease presence by the unnatural action of speaking to an invisible person [22]. There has been a lack of research regarding the effects of using intradiegetic $_{IN}$VR-Qs for VR research studies.

Contributions

In this paper, we want to extend the works by Schwind et al. [35] and Putze et al. [30] by proposing that embedding an evaluation fully into the experienced VE, thus using intradiegetic $_{IN}$VR-Qs, would mitigate some effects of both post-test questionnaires and extradiegetic $_{IN}$VR-Qs, and are preferred by experts and users alike. Based on related works, we designed four concepts of $_{IN}$VR-Qs that vary in their degree of being integrated into the VE, ranging from a complete extradiegetic to fully intradiegetic design 1. In a first study, experts compare these UIs with each other regarding UX, presence, usability and preference. In the second study, we explore various effects of $_{IN}$VR-Qs, represented by the preferred concept of the first study, and survey users' opinions about this evaluation method in comparison to the one composed in reality that we call *$_{OUT}$VR-Qs*.

We contribute by investigating the following hypotheses:

H1 Intradiegetic UIs are preferred over extradiegetic UIs.
H2 $_{IN}$VR-Qs are majorily preferred over $_{OUT}$VR-Qs.
H3 Intradiegetic $_{IN}$VR-Qs do not affect presence.
H4 $_{IN}$VR-Qs take less time compared to $_{OUT}$VR-Qs.

Our work will demonstrate advantages of including evaluations directly in the experienced VE. Our findings focus on research scenarios of the automobile industry but are important for any other field conducting research in VR and augmented reality (AR), as well. We will also emphasize the importance of conducting further research on the effects of $_{IN}$VR-Qs in general and especially $_{IN}$VR-Qs that are designed as intradiegetic UIs.

2 RELATED WORK

There is already a large body of related works studying various VR-related aspects. The concept of presence and the usage of VR-questionnaires are the most relevant for this paper and thus will be shortly discussed. This section will also show how we advance with the state of the art.

2.1 Presence

Presence is a widely researched aspect. In most cases, presence is understood as the subjective *sense-of-being-there* in a VE [37, 39, 44]. The creation of this cognitive state is regarded as *the* essential feature of a VE [35] and thus can be seen as a measurement for the VE's quality [37].

Although presence is highly subjective, technical features and some aspects of the created VE can minimize the possibility of so-called breaks-in-presence (BIPs) [39]. Field-of-view, resolution and correct colour representation are constantly ameliorated in Head-Mounted-Displays (HMDs) [25], correct application of physical laws, environmental sound and haptic feedback enhance the VE's realism [25, 35], and being represented with a virtual body in VR all enhance the possibility for users to feel present [20, 25]. .

Presence can be measured via several physiological parameters [25, 39], behavioral assessment [30] or by counting the amount and intensity of BIPs [26, 39]. The predominant method is self-reporting in post-test presence questionnaires [37]. Questionnaires are shown to be sensitive enough to reveal differences in presence [21] and have the benefit of being easily administerable and comparable [21, 37]. However, criticism of this method includes reliability and validity aspects [26], not being sensitive to specific events [26], and, for the most commonly used presence questionnaire by Witmer and Singer [44], its subjectively defined factors and few items directly assessing presence [33, 38].

2.2 VR Questionnaires

Post-test questionnaires, here named $_{OUT}$VR-Qs, have several negative effects, mainly the need to remember the experienced scene [15], demotivation [17] and loosing the representation of spontaneous thoughts in the raw data [13, 17]. Several findings suggest the mitigation of those by embedding presence questionnaires into VR, thus designing $_{IN}$VR-Qs [30, 35]. Alexandrovsky et al. [3] verified the need for such an evaluation by conducting an expert survey in which 64% stressed the importance of $_{IN}$VR-Qs. 82% would, in general, prefer $_{IN}$VR-Qs over $_{OUT}$VR-Qs.

$_{IN}$VR-Qs are designed as user interfaces (UIs). UIs are interactive elements that have the task to enhance the dialog between machine and human by transferring information [22]. All concepts of $_{IN}$VR-Qs can be classified into either being *extradiegetic* or *intradiegetic* UIs. In line with the official definition of the game engine Unity3d [40], we use extradiegetic to qualify that additional information is detached from the VE's context, e.g. by presenting them in external menues. Intradiegetic UIs become part of the actual VE, e.g. by utilising already implemented monitors of the VR-scene to show relevant information [40]. The more the four concepts are embedded into the VE's content, the more they qualify for intradiegetic UIs [15, 40].

Based on literature research of Alexandrovsky et al. [3], fifteen papers published in the five major digital libraries used some form of $_{IN}$VR-Qs in the years from 2016 to 2019[1]. Most of them use extradiegetic UIs. Their design choices differ from each other in (1) the means of presentation, for example using floating head-up-displays [9, 12, 23], world-referenced UIs [14, 18, 24, 29, 34–36] or UIs attached to the body [43], (2) the means of interaction, mainly by using VR-controllers [18, 23, 24, 29, 35], gamepads [2, 9, 14], answering orally [12] or using free-hand interaction via Leap Motion gesture-based inputs [32, 34, 36, 43], and (3) other design choices like presenting single or multi-questions at once, answering via Likert-scales [29, 34–36], text-based buttons [23] or sliders [2, 9, 14, 18].

In some experimental setups, the VR-scene (mostly a game) was replaced entirely by a new VE, either by showing the questionnaire in front of a plain background [9, 23] or in a virtual replica of the laboratory [19, 30, 35]. This approach resembles the *UI Extern* used in this paper (see figure 1 a). Other studies show the questionnaire as a paper-based, mostly 2D, replica overlaying the VE [14, 29, 34]. Our concept *UI Canvas*, which is attached to the HMD, correlates with this design choice (see figure 1b).

Only few studies use $_{IN}$VR-Qs designed as intradiegetic UIs. Wienrich et al. [43], for example, attached the questionnaire to the virtual hand of the user, which can be broadly defined as a virtual object of the experienced VE. In this paper, we use the *UI 3D-font* to resemble UIs that are partly included in VEs by being attached to a virtual object (see figure 1c). None of the above mentioned VR-studies use already existing structures of the experienced VE to display $_{IN}$VR-Qs. However, this concept was successfully tested in a PC-setting by Frommel et al. [15]. With the concept of the *UI Tablet* we transfer this last approach to VR (see figure 1d).

Five studies utilise $_{IN}$VR-Qs to evaluate presence [16, 30, 34–36]. Only two investigate the effects of $_{IN}$VR-Qs on presence in contrast to a classical post-test evaluation conducted in reality that we call $_{OUT}$VR-Qs [30, 35]. Schwind et al. [35] proved that three standard questionnaires for presence measurement provide similar results in VR as in reality. Thus, a transferability of questionnaires to the medium VR is in principle possible. Based on these findings, Putze et al. [30] compared the occurrence of BIPs when using $_{IN}$VR-Qs and $_{OUT}$VR-Qs. Their study indicates that $_{IN}$VR-Qs reduce BIPs and systematic bias while the player experience is not negatively affected.

Schwind et al. [35] assumed that by switching the VE with another, thus using extradiegetic UIs, they still recreated similar effects as in $_{OUT}$VR-Qs, for example the need to remember the experienced scene. Thus, it is suggested to stay close to the context of the ongoing experience [15], which was successfully tested regarding 2D PC settings [15], and broadly include other attempts like the thinking-aloud protocol, in which the participant permanently speaks his thoughts aloud [25], and several modes of verbal interviewing, as well [5, 10]. However, the unnatural action of speaking to an invisible person can in turn decrease presence [22].

2.3 Summary

Previous research highlights the importance of presence as a measurement of a VE's quality [35, 37]. Despite different approaches to measure presence [25, 26, 39], the predominant method relies on using questionnaires [3, 21, 30, 37]. Further related work emphasize the benefits of integrating these questionnaires into VR [3, 30, 35]. Fifteen approaches of $_{IN}$VR-Qs already used in literature have been clustered based on design choices, which forms the basis for the creation of four types of $_{IN}$VR-Qs used in our first study 1. However, related works shows a lack of research regarding the effects of $_{IN}$VR-Qs in general and especially of $_{IN}$VR-Qs designed as *intradiegetic* UIs.

3 STUDY 1: CONCEPTION OF $_{IN}$VR-QS

In the first study, we explored the acceptance of four UI concepts of how a questionnaire could be integrated in VEs (see figure 1). Besides investigating which type of UI (extradiegetic or intradiegetic) is preferred (**H1**), we also determine the most preferred UI for presenting $_{IN}$VR-Qs for our specific study.

The study used a within-subject design. This approach enables the participants to decide on their preferred evaluation method after having experienced all concepts. The study employed a single factor variance. We investigated the independent variable *concept* that exhibited the four factor levels *UI Extern, UI Canvas, UI 3D-font* and *UI Tablet*, presented in figure 1. Because of possible sequence effects, the order of presented design concepts was varied and equally randomized between participants. The conditions were counterbalanced.

3.1 Participants

We conducted the study with $N = 11$ participants (5 male and 6 female) with an average age of 26.32 (min: 20, max: 28). They work for the German automotive company *Volkswagen AG*. All of them had previous experience in using VR systems (31% work with VR on a daily basis). 21% of the participants additionally work in the design sector, 32% test the usability of products and 16% conduct studies as their regular content of work. This mixture of experts promises comprehensive insights of $_{IN}$VR-Qs from different perspectives, focusing on strengths and weaknesses of the UI and taking typical

[1]The work of Haas [19] does not appear in this list for it was only published internally at a German university. The research of Putze et al. [30] was not included because it was published in 2020.

Figure 2: Study environment for study 1: Left picture shows view from outside (participant sits on a car seat), right picture shows the VE seen by participant.

user behaviour in a study into account. They took part in the study without any further compensation but within their regular working time.

3.2 Apparatus

An HTC Vive Pro with 1200x1600 resolution per eye was used as a Head-Mounted-Display (HMD). We used a Leap Motion sensor attached to the front of the HMD to convert the users hands' movements into VR. The Leap Motion standard hand model was used. A Desktop-PC with a NVIDIA GeForce GTX 1080 Ti video card, an Intel® Core™ i7-2600 CPU with 3.40GHz and 24GB RAM was used to create and run the VR simulation around 90 frames per second.

3.3 Study Environment

For this study, a car seat was placed in the center of a 4x4m tracking space (see figure 2). For the VE, we reproduced a room used for driving simulations in the automobile industry that contains a minimal representation of a car's cockpit [8]. This setting was enriched with three cubes that could be stacked, as well as a Volkswagen-Logo in the center of the steering wheel that changes color when touched (see figure 2).

In order to verify the first hypothesis **H1**, we designed four UIs of $_{\text{IN}}$VR-Qs (see figure 1) that reflect approaches of $_{\text{IN}}$VR-Qs featured in related works (see section 2.2). Each UI features a different level of integration into the VE: *UI Extern* is completely removed from the ongoing experience, *UI Canvas* is attached to the HMD and overlays the VE, *UI 3D-font* consists of 3D-letters attached to the virtual steering wheel and *UI Tablet* is embedded as much as possible into the VE by being located on a monitor already existing in the VE. Two design concepts (UIs a and b) can be classified as *extradiegetic* UIs [40], as they overlay the actual VE and present information outside the VE's context. UIs c and d fit themselves into the VE and thus represent intradiegetic UIs [15, 40].

Each UI consists of two parts. The first is a questionnaire block displaying several statements, for example "The steering wheel was very realistic." or "I could move the cubes intuitively." The second, the assessment block, is designed as a five-point-Likert scale to judge one's agreement with the statements. An input can be made by touching the right box in the assessment block with the virtual replica of the Leap Motion tracked hands. The order of these two elements varies depending on different layouts of UIs.

3.4 Procedure & Measures

After a formal introduction, in which the general procedure was explained, the participants took a seat in the car's cockpit 2, put on the HMD and adjusted the interpupillar distance. As all participants had experience with VR systems, we did not need to test for stereoscopic viewing abilities.

As a first task, read aloud by the examiner, the participants changed the colour of the steering wheel by touching it. Using one of the $_{\text{IN}}$VR-Qs-concepts, the participants gave their opinion about the realism of the scene while still having the HMD attached. Afterwards, they fulfilled a second task, in which the participants stacked three cubes on top of each other. They gave their feedback on the simplicity of the interaction with help of the same concept of $_{\text{IN}}$VR-Qs as before. With this, the participants finished the first round.

Afterwards, the participants took off the HMD and filled out a custom post-test questionnaire. We decided against using standardized questionnaires because they are not able to inquire aspects adapted to the specific study environment needed. It features six specific questions regarding user experience (UX) and presence each, as well as four about usability aspects. All were answered via a six-point-Likert scale. The subjects further stated the length of VR-sessions for which they could imagine using the experienced UI. They then graded the concept (best = 1, worst = 6) and gave further qualitative feedback.

As we chose a within-subjects design in order to better evaluate the differences in the UIs, all participants experienced all four to be assessed UIs in a randomised order determined by a latin square. Hence, the described procedure, including tasks, $_{\text{IN}}$VR-Qs and post-test evaluation about the UI was performed four times in total per participant with a different UI each. In a final post-test evaluation, the participants compared all four concepts, ranked them, and offered further qualitative feedback about improvement possibilities in a semi-structured interview. The interviews were transcribed and compared via a qualitative content analysis [28].

3.5 Results

The data was analysed using Friedman's ANOVA for repeated measures and the Wilcoxon signed-rank test for paired samples with Bonferonni correction as post-hoc test. It was measured at a .05 confidence level. The data appeared to have a good internal consistency, calculating Cronbach's Alpha regarding UX ($\alpha = .866$), presence ($\alpha = .802$), and usability ($\alpha = .814$).

Figure 3 shows the results regarding UX, presence and usability collected after each UI test. The maximum possible value for UX and presence is 36 and for usability 24. We found significant differences ($\chi^2(11) = 88.028, p < .0001$). Regarding the aspect of UX, we found a significant difference between *UI Extern* and both *UI 3D-font* ($p = .008$) and *UI Tablet* ($p = .004$), as well as between *UI Canvas* with UI 3D ($p = .01$) and *UI Tablet* ($p = .004$). Similar results were found concerning presence: *UI Extern* differs significantly from *UI 3D-font* ($p = .004$) and *UI Tablet* ($p = .004$), and *UI Canvas* from UI 3D ($p = .004$) and UI Tablet ($p = .004$), as well. Usability shows significant differences between *UI Canvas* and both intradiegetic concepts, *UI 3D-font* ($p = .033$) and *UI Tablet* ($p = .041$). In all cases, the intradiegetic UIs received better results.

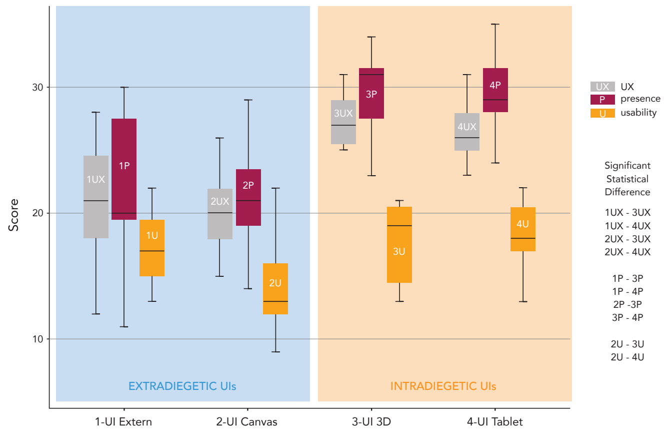

Figure 3: UX, presence and usability scores of the first study. Extradiegetic UIs appear on a blue, intradiegetic UIs on an orange background. Statistically significant results are marked in the legend.

We also analysed the grades assigned to each concept (best = 1, worst = 6). The *UI Extern* received a grade of $M = 3.36$ ($SD = 0.81$), *UI Canvas* $M = 3.91$ ($SD = 1.38$), *UI 3D-font* $M = 2.55$ ($SD = 1.04$) and *UI Tablet* also $M = 2.55$ ($SD = 0.82$). The results regarding the length of VR-sessions in which the UIs are imaginable are not statistically significant. Nevertheless, we found a tendency towards using extradiegetic UIs only in VR-sessions up to 15 min due to disorientation caused by the change of VEs. Intradiegetic UIs were imaginable for all lengths of VR-sessions.

In the overall ranking of all UIs, 91% preferred one of the intradiegetic concepts. 45% specified that the reason for this choice was the full embedding of intradiegetic UIs in the VE. They enjoyed still being able to see and interact with the VR-scene while filling out questionnaires. 64% preferred the *UI Tablet* above all else. Of those, 57% stated that the black background helped to estimate the distance between virtual finger and real tablet surface. This facilitates the usability. Additionally, 57% specified that the *UI Tablet* was the least intrusive UI and is easy transferable to other VR-settings.

4 STUDY 2: EFFECTS OF $_{IN}$VR-QS

In the second study, we investigated the effect of $_{IN}$VR-Qs in comparison to classical post-test evaluations conducted in reality. We assessed which evaluation method is preferred (**H2**) and which effects $_{IN}$VR-Qs have on presence (**H3**) and questionnaire completion time (**H4**). Based on the findings of the first study, we utilised the fully intradiegetic evaluation concept *UI Tablet* as $_{IN}$VR-Qs. We improved this prototype based on the feedback gained in the qualitative expert survey to increase its usability and visual appearance, more navigational opportunities and audio-feedback when a check-box is selected. It accommodates the now-used 7-point-Likert scale of the standardized presence questionnaire by Witmer and Singer [45] displayed on it. As we provide haptic feedback in study two (see section 4.2), the size of *UI Tablet* is limited to 12.9" to fit the screen of a Microsoft surface tablet.

Figure 4: Study environment of study 2: The left photograph shows the real laboratory setup, the right image depicts the virtual replica.

This study deployed a single factor variance within-subject design with repeated measures. As an independent variable, the *test condition* was tested in two factor levels (1) scene in VR and evaluation in reality shortened to *VR_real* and (2) scene in VR and evaluation in VR as well, shortened to *VR_VR*. The order of the experienced test conditions was varied and allocated to the participants in a randomised manner to avoid a possible sequence effect. The conditions were counterbalanced.

4.1 Participants

This study incorporated $N = 24$ participants (15 male and 9 female) with an average age of 31.73 (min: 21, max: 53). All participants were inexperienced with VR, as 78% have made individual experiences with VR once a year or less. The participants on average scored 4.3 of 6 points in the *Access Affinity for Technology Interaction* scale (ATI-questionnaire) and thus can be considered to be technically proficient. All users participated voluntarily and were compensated with a 20€-gift card. In total, the study took one and a half hours per participant.

4.2 Apparatus

In contrast to study 1, we used a HP Reverb (first generation) as HMD due to the better resolution (2160x2160 pixel per eye) and wider field-of-view (114°) compared to the HTC Vive Pro [11]. The system was driven by an NVIDIA Quadro P6000 videocard, two Intel® Xeon® CPUs E5-2667v4 with 3.2GHz each and 128GB RAM. We used the standard Leap Motion hand model wearing gloves to prevent skin bias. The hands were not calibrated to the actual users hand. We decided against a whole-body avatar due to the *uncanny valley* effect [25].

4.3 Study Environment

For the study environment, we again used the virtual representation of a room used for driving simulations [8]. However, this time the participants took place in the car's cockpit and entered a VE enhanced by haptic feedback (see figure 4). Haptic feedback was established through a registration of the virtual dashboard to the physical object by manually shifting the room model into an as accurate as possible position and orientation. By means of a tracked Windows Mixed Reality controller, the movement of the steering wheel was translated into the correct rotation in VR. This approach offers the chance to enhance the realism of the scene and supports

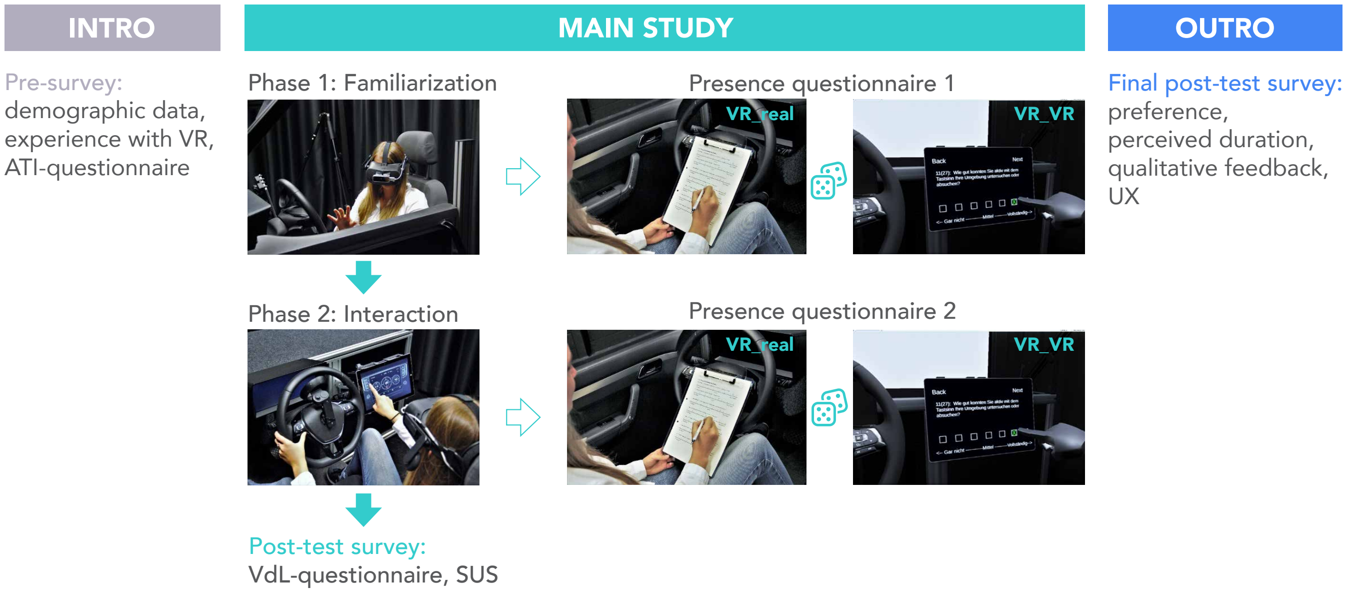

Figure 5: Procedure of study 2: Intro includes pre-survey. Main study consists of a familiarization and interaction phase, respectively followed by presence questionnaires conducted in reality (VR_real) or in VR (VR_VR), experienced in a randomised but counterbalanced order. A post-test survey evaluates acceptance and usability. Outro includes final post-test survey.

an intuitive interaction which in turn increases the chance of higher presence [25].

A tablet was installed to the right of the steering wheel. We utilised it for two reasons. First, we streamed a fully functional real car's infotainment module on it. This is a system typically located on the front board that allows the user to access navigation and media controls in a car. Second, the tablet offered haptic feedback for the *UI Tablet* on which the presence $_{IN}$VR-Qs is displayed. The position and orientation of the physical tablet was registered to the virtual tablet by the same procedure as done with the dashboard. All registration procedures were done each time before participants joined the study.

4.4 Procedure & Measures

The procedure is presented in figure 5. First, the participants were welcomed in an anteroom and were tested for their interpupillar distance and stereoscopic viewing abilities. For this purpose, the participants wore polarisation glasses and identified different deep patterns to determine their level of depth perception. All participants scoring 80% or more were then told the general procedure. They filled out a pre-test-questionnaire about demographics, including a section about previous experience in VR, and the ATI-questionnaire.

In order to investigate the preferred evaluation method ($_{IN}$VR-Qs vs $_{OUT}$VR-Qs) we designed two test conditions (VR_real and VR_VR) that differ in the medium of conducting the evaluation (see figure 5). As we chose a within-subject design in order to better compare the two conditions, all participants experienced both test conditions in randomised order. The conditions were counterbalanced.

In test condition VR_real, the subjects entered the laboratory, took a seat in the driver's seat and put on the HMD to enter the VR-replica of the laboratory (see figure 4). The participants then experienced the first phase, a familiarization period for about two minutes in order to adjust to the scene (see figure 5). Then, they filled out a first presence questionnaire via pen-and-paper (VR_real-1). Afterwards, the participants moved on to the interaction phase. The examiner read aloud ten tasks that were completed by the participants using the infotainment system presented on the virtual tablet. The tasks tested skills in media usage (e.g. "Please turn on the traffic broadcast") and navigation (e.g. "Please navigate to the nearest gas station"). After the interaction phase, the subjects filled out a second presence questionnaire (VR_real-2). At the end of test condition VR_real, the participants left the whole laboratory setting and filled out a post-test questionnaire. Content of the post-test survey was to evaluate the evaluation itself, not the interaction with the tasks. We used diverse standardized questionnaires. The Van der Laan Acceptance questionnaire (VdL) was utilized to measure the amount of acceptance and the System-Usability-Scale (SUS) to survey the usability of the experienced evaluation method of the test condition.

In test condition VR_VR, the two phases (familiarization phase and interaction phase) remain the same as in condition VR_real. However, this time the presence questionnaires were filled out in VR (VR_VR-1 and VR_VR-2). For both presence evaluations, the *UI Tablet* was utilised as $_{IN}$VR-Qs.

After both test conditions have been experienced, the participants completed a final questionnaire. It included questions about the preferred evaluation method, the perceived duration of both test conditions and UX-aspects. They also had the possibility to give qualitative feedback.

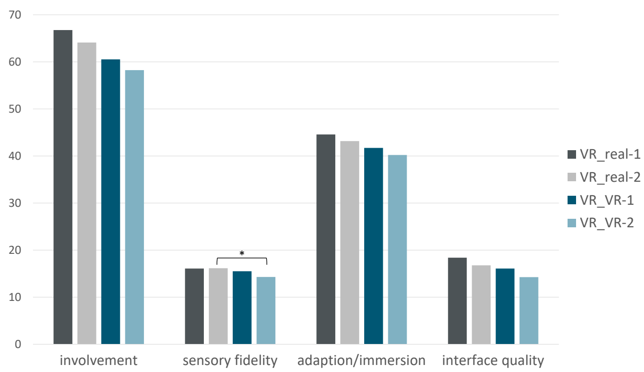

Figure 6: Results for presence subscales for each presence measurement. Statistically significant results ($p \leq .05$) are marked with *.

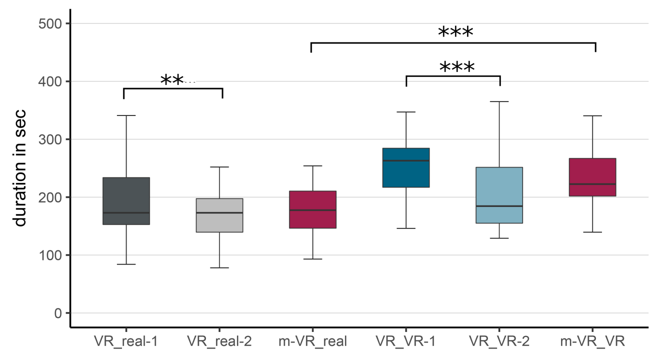

Figure 7: Questionnaire completion time for all presence measurements. The calculated means of each test condition are shown in red. Statistically significant results ($p \leq .01$) are marked with **, results $p \leq .001$ with *.**

In both test conditions, we decided to deploy the presence questionnaire of Witmer and Singer in the version of 2005[2] [45]. Although being aware of criticism (see section 2.1), it is still one of the most often used presence questionnaires [3], which has been proven to lead to similar results when embedded into VR [35]. In comparison to other presence questionnaires, it leads to the least variance in the results [35]. Due to the static nature of the VE and the user remaining seated in VR, drastic changes in VR-sickness, for example nausea, are not to be expected and were thus not measured.

4.5 Results

The data was analysed using Friedman's ANOVA for repeated measures and the Wilcoxon signed-rank test for paired samples with Bonferonni correction as post-hoc test. It was measured on a .05 confidence level.

4.5.1 Presence. The presence questionnaire of Witmer and Singer [45] structures presence in four sub-categories: involvement, sensory fidelity, adaption/immersion, and interface quality. Our results in that regard are presented in figure 6. First, we compared first and second presence measurement within each test condition (VR_real-1 with VR_real-2, and VR_VR-1 with VR_VR-2) and found no significant differences. Second, we compared first and second measurement between the conditions (VR_real-1 with VR_VR-1, and VR_real-2 with VR_VR-2). Only sensory fidelity was evaluated significantly lower in VR_VR-2 compared to VR_real-2, $\chi^2(3) = 9.15$, $p = .027$, $r = -.53$, $Z = -2.608$.

4.5.2 Questionnaire Completion Time. We further surveyed the duration of the presence evaluations (see figure 7). In both test conditions (VR_real and VR_VR), the first measurement took significantly longer (VR_real: $p = .009$, $r = -.53$, $Z = -2.615$; VR_VR: $p = .001$, $r = -.69$, $Z = -3.386$). In VR_real, the mean for VR_real-1 was $M = 187.17sec$ ($SD = 60.49$) and for VR_real-2 $M = 169.88sec$ ($SD = 51.71$). Thus, the second presence measurement was completed $17.29sec$ faster than the first. In condition VR_VR, the mean for VR_VR-1 was $M = 270.08sec$ ($SD = 94.49$) and for VR_VR-2 $M = 211.38sec$ ($SD = 66.92$). So, in VR_VR the second presence measurement was $58.7sec$ faster than the first.

Further, we calculated an average duration for both presence measurements per condition, presented in red in figure 7. For the test condition VR_real participants needed $M = 178.52sec$ ($SD = 54.37$) on average, and for condition VR_VR $M = 240.73sec$ ($SD72.27$). This corresponds to a difference of $62,21sec$ that $_{\text{IN}}$VR-Qs take longer to complete. The results were significant, $p < .0001$, $r = -.88$, $Z = -3.745$.

4.5.3 Acceptance. The acceptance of both evaluation methods was measured with help of the VdL-questionnaire [41]. We found no significant difference for satisfaction, $p = .626$, and usefuleness, $p = .140$.

4.5.4 Usability. Usability was assessed with the SUS [7]. We could not find any significant difference between the test conditions, $p = .058$.

4.5.5 Preference. Our findings regarding the preferred evaluation method are shown in figure 8. 67% of the participants would prefer $_{\text{IN}}$VR-Qs. Moreover, 83% report a more joyful experience when filling out $_{\text{IN}}$VR-Qs. However, the participants also thought $_{\text{IN}}$VR-Qs had less usability (75%), were slower (79%) and induced a feeling of irritation (67%).

We also asked the participants to give reasons for their choices in the post-test survey. The answers were categorised and evaluated with help of a frequency analysis. 48% of all comments were related to usability aspects. Of those, problems with selecting the right check box when using $_{\text{IN}}$VR-Qs rose up in 62% of the cases and 30% relate to tracking problems of the hands. 13% of further comments relate to being glad not needing to remove the HMD for the evaluation process. 11% apply to $_{\text{IN}}$VR-Qs being perceived as a more modern, environmental friendly and motivational approach than $_{\text{OUT}}$VR-Qs. 10% explicitly noted the high realism of the VE. The remaining 18% split themselves up into comments about the experimental setup and VR in general.

[2]Due to the lack of auditory aspects in our VE, we removed question items 5, 11 and 12 from the original presence questionnaire.

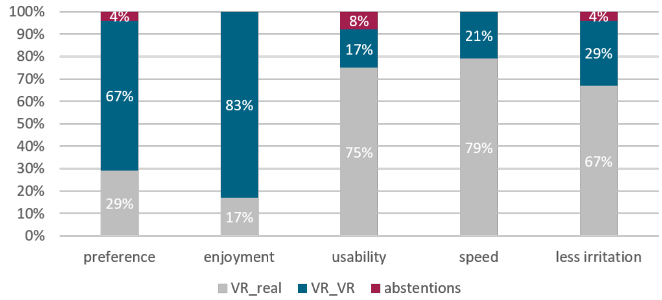

Figure 8: Results of the post-test evaluation regarding the preferred evaluation method in percent. The participants decided between VR_real (grey) and VR_VR (blue) for each aspect separately.

5 DISCUSSION

In our studies, we compared extra- and intradiegetic design concepts of $_{\text{IN}}$VR-Qs and conducted a user study to assess the effects of $_{\text{IN}}$VR-Qs in comparison to $_{\text{OUT}}$VR-Qs. Our outcome and limitations will be discussed in the following sections.

5.1 Preference

In the first study, we compared four different design concepts of $_{\text{IN}}$VR-Qs with each other(see figure 1). The results show that intradiegetic concepts are significantly favoured, supporting **H1**. As stated in the qualitative feedback, the main reason for preferring intradiegetic concepts was being able to see and interact with the VE while filling out questionnaires, and the evaluation process felt the least intrusive for many participants.

In the second study, we compared $_{\text{IN}}$VR-Qs with classical post-test evaluations conducted in reality. In the expert survey of Alexandrovsky et al. [3] it turned out that 82% of the VR-experts preferred $_{\text{IN}}$VR-Qs over $_{\text{OUT}}$VR-Qs. This insight matches our findings although we surveyed VR-laymen. 67% of our participants preferred the experienced $_{\text{IN}}$VR-Qs. This supports our hypothesis **H2**. Moreover, 83% report a more joyful experience when filling out $_{\text{IN}}$VR-Qs in our study. These results are comparable to findings of Alexandrovsky et al. [3], whose qualitative feedback also proved higher enjoyment for $_{\text{IN}}$VR-Qs. In our study, we reached a high preference although $_{\text{IN}}$VR-Qs were perceived less usable, slower and more irritable. Based on the qualitative feedback, we can assume that these feelings correlate with the error-prone usability. Reasons for that are discussed in section 5.4. The high preference of $_{\text{IN}}$VR-Qs in our study might also be due to the novelty aspect of $_{\text{IN}}$VR-Qs. However, the qualitative feedback of our two studies combined with the findings of Alexandrovsky et al. [3] lead to the conclusion that $_{\text{IN}}$VR-Qs are in general experienced as being more joyful and helpful due to their integration in VR.

5.2 Presence

Our results show that presence is unaffected by the form of evaluation method (see figure 6), thus supporting our hypothesis **H3**. These results are in accordance with study results of Schwind et al. [35]. However, the values show a decrease between the first and second measurement per condition. This is likely due to the high focus on the car's infotainment module streamed on the tablet in the interaction phase. Due to the high complexity of the content, participants experienced high latency and worse resolution compared to the rest of the environment. In the qualitative feedback, 17% of the participants stated that the bad experience of interacting with the infotainment module directly influenced the presence evaluation. This is also likely the reason for the statistically significant results in the subscale sensory fidelity, which describe how well participants are able to sense different aspects of the VE [45]. In combination with some technical limitations elaborated in section 5.4, this also explains why the presence values are the lowest in measurement VR_VR-2. We also felt as if the Witmer and Singer presence questionnaire [45] might not have been able to depict presence as detailed as necessary for our study, following some criticism in literature about its subjectivity and length [35, 38]. Presence might also in general be better measured using other means like counting BIPs or physiological measures 2.1. Using one of these methods or other presence questionnaires could be enlightening and should be researched in future studies.

5.3 Questionnaire Completion Time

Against our assumption **H4** and in contrast to findings of [19, 30, 35], we found significant differences for the duration of completing questionnaires in both conditions.

First, our results show that in both conditions the second presence questionnaire took significantly less time than the first. This could be explained by getting familiar with the questions, thus answering faster. However, in the condition VR_VR the participants were close to $one min$ faster in the second measurement while the time difference in VR_real only accounted for $17 sec$. Thus, we also assume a training effect regarding the answering via $_{\text{IN}}$VR-Qs. We expect that a prior training with the *UI Tablet* would have reduced the measured time for $_{\text{IN}}$VR-Qs in our study and should be tested in future research.

Second, our results suggest that $_{\text{IN}}$VR-Qs take significantly more time than $_{\text{OUT}}$VR-Qs. Apart from the training effect, we assume inaccuracies between VR and reality due to technical issues elaborated further in the limitations (see section 5.4) to be the reason. Incorrect entries could happen, which needed to be fixed manually by pressing the back button, which took time. Our findings are, thus, in line with other studies that also showed lower usability for $_{\text{IN}}$VR-Qs [3]. Another explanation might be the higher engagement with the question's content. For example, they interacted with the steering wheel again after reading the question "How well could you move or manipulate objects in the virtual environment?". We experienced such actions with more than half of the participants. This only happened in condition VR_VR, though, due to being able to still see and interact with the VE. As a downside, initial thoughts of participants might get lost due to having more time to re-think each question. We nevertheless strongly assume that the results of the questionnaire using $_{\text{IN}}$VR-Qs were more reliable because the participants did not rely solely on their memory and gut feeling anymore.

As a last thought, we want to emphasize that it takes time for participants to re-orient themselves in the new environment after an HMD was removed and put back on. This assumption is based on findings of Schwind et al. [35]. Unfortunately, due to technical problems (see section 5.4), we were not able to cleanly measure this time difference consisting of the duration of the evaluation (that we could measure) plus the time to reorient in the new environment (that we could not measure). Therefore, the questionnaire completion time for the condition VR_real could be actually higher than indicated in this study.

5.4 Limitations

The usability measured with the SUS was not statistically different between the two test conditions ($p = .058$). However, the evaluation of the qualitative feedback showed that usability was the most often mentioned issue. Some participants stated that they involuntarily checked wrong boxes which needed to be corrected manually by using the provided back-button. During the study, we found several technical issues that correlate with this problem. First, we experienced strong heat development in the Leap Motion sensor attached to the front of the HMD, which might have been also influenced by an increase in temperature of the used HMD. This could have led to an incorrect tracking of the fingers.

Second, we found that sometimes the VE did not exactly fit to the haptic feedback provided by the real study environment although we tried to superimpose them on each other as precisely as possible. This could be explained by the markerless inside-out-tracking of the HMD, the HP Reverb [11]. We experienced that especially after removing and donning the HMD again, it recalculated its position, resulting in inaccuracies. As we could not guarantee a perfect calibration of VE and real surrounding, we applied an input tolerance of $.3mm$ beforehand.

Third, we experienced challenging imprecise tracking of the hands in interaction [1]. Because the *UI Tablet* was attached to the right of the dashboard, the users could only interact with nearly outstretched arms (about $50cm$) and outside the optimal tracking space directly in front of the sensor [27], which decreased the sensor's precision and ability to correctly identify the fingers [1]. Some participants additionally had long fingernails that tricked the sensor to calculate the virtual hands wrongly closer to the tablet. Some participants also used their fingers to read the questions on the tablet. Due to the input tolerance explained above, sometimes this behavior led to an involuntarily checking of boxes.

Due to these insights, it would be interesting to see changes in the results if another form of input than the Leap Motion tracked hands is used. In our concrete setup, this option was second to the benefits of a natural interaction enriched with haptic feedback, but will be subject of future research.

We are aware that our approach of $_{\text{IN}}$VR-Qs is mostly applicable to our specific setting, in which an already existing monitor of the VE could be easily re-used to display the questionnaire. Equivalents need to be find for each experience individually [15]. Although that might be especially difficult for outdoor VEs (where we could imagine a questionnaire being presented on a signpost or a stone) or embodiment scenarios (in some settings [31] we could imagine a questionnaire being integrated on a sticky note attached to the mirror)

the benefits of including $_{\text{IN}}$VR-Qs should be carefully weighed each occasion separately. In some cases, embedding a questionnaire within a VE might feel more disturbing to the user than relying on verbal administration [5] or the thinking-aloud protocol [25]. Although literature suggests that these methods can disturb presence [22], and because the mode of questionnaire administration is likely to affect data quality [5], a comparisons between verbal methods and $_{\text{IN}}$VR-Qs should be conducted in future research. Besides the restrictions discussed above, we see $_{\text{IN}}$VR-Qs generally valuable for research conducted in VR and, as an extension, also for AR. Additionally, our results suggest a general preference for in the experienced environment integrated questionnaires whatever the research method. Thus, our findings suggest relevance for all research methods [42], including online, lab and in-situ studies, which should be further researched in the future.

6 CONCLUSION

The evaluation of products or systems in VR is mostly conducted in the form of post-test questionnaires. To mitigate the emerging memory effects, inaccuracies in the evaluation and BIPs, questionnaires can be included directly in the VE [3, 15, 19, 30, 35]. However, a literature analysis shows that little research has been conducted to fully comprehend the effects of $_{\text{IN}}$VR-Qs [3, 30, 35].

We contribute to this field of research with two studies: In the first study, we tested four design concepts of $_{\text{IN}}$VR-Qs with varying level of being embedded in a VE. We found that intradiegetic UIs guarantee a higher UX and are widely preferred by experts, confirming **H1**.

In the second study, we used the best rated concept (*UI Tablet*) as representative for $_{\text{IN}}$VR-Qs. In a user study ($N = 24$), we compared $_{\text{IN}}$VR-Qs with classical $_{\text{OUT}}$VR-Qs regarding preference (**H2**), effects on presence (**H3**), and completion time (**H4**). 64% of the users preferred $_{\text{IN}}$VR-Qs over $_{\text{OUT}}$VR-Qs, confirming **H2**, because of a high UX, not needing to remove the HMD and the full integration of the UI in the VE. Presence was neither enhanced nor reduced by $_{\text{IN}}$VR-Qs, which supports **H3**. Against our previous assumption, the duration of completing questionnaires in VR increased significantly, denying **H4**. This is probably due to some technical issues, especially regarding registration and tracking quality of HMD and Leap Motion sensor. Nevertheless, we also experienced a higher interactive engagement with the questions' content that suggest an increase of accuracy in the questionnaire. To measure this effect in detail, further studies should be conducted.

In total and with some limitations discussed in 5.4, our findings demonstrate the potential of $_{\text{IN}}$VR-Qs and we suggest using fully integrated VR-questionnaires, that we refer to as *intradiegetic $_{IN}$VR-Qs*, as evaluation method in future VR studies. We recommend using virtual screens embedded in the VE to display a presence questionnaire.

Further research could measure the effect of prior training, the implementation of a comment function desired by some participants, and other interaction forms than gesture-based inputs to mitigate technical challenged experienced in both our studies. Additionally, it should be explored to what extent our findings about $_{\text{IN}}$VR-Qs apply to both usability testings and other research methods [42].

ACKNOWLEDGMENTS

We thank all participants for taking part in the studies. The involved organizations funded the studies internally.

REFERENCES

[1] A. Ahmad, C. Migniot, and A. Dipanda. 2017. Tracking Hands in Interaction with Objects: A Review. In *2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*. IEEE, Jaipur, 360–369.

[2] Majed Al Zayer, Isayas B. Adhanom, Paul MacNeilage, and Eelke Folmer. 2019. The Effect of Field-of-View Restriction on Sex Bias in VR Sickness and Spatial Navigation Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300584

[3] Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan David Smeddinck. 2020. Examining Design Choices of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3313831.3376260

[4] Leif Berg and Judy Vance. 2016. Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Reality* 21, 1 (09 2016). https://doi.org/10.1007/s10055-016-0293-9

[5] Ann Bowling. 2005. Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health* 27, 3 (05 2005), 281–291. https://doi.org/10.1093/pubmed/fdi031 arXiv:https://academic.oup.com/jpubhealth/article-pdf/27/3/281/4347729/fdi031.pdf

[6] Jennifer Brade, Mario Lorenz, Marc Busch, Niels Hammer, Manfred Tscheligi, and Philipp Klimant. 2017. Being There Again Presence in Real and Virtual Environments and Its Relation to Usability and User Experience Using a Mobile Navigation Task. *Int. J. Hum.-Comput. Stud.* 101, C (May 2017), 76–87. https://doi.org/10.1016/j.ijhcs.2017.01.004

[7] John Brooke. 1996. SUS-A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*, P.W. Jordan, B. Thomas, B.A. Weerdmeester, and A.L. McClelland (Eds.). Taylor and Francis, London, 189–194. Issue 194.

[8] Heiner Bubb, Klaus Bengler, Rainer E. Grünen, and Mark Vollrath. 2015. *Automobilergonomie*. Springer Fachmedien, Wiesbaden. https://doi.org/10.1007/978-3-8348-2297-0

[9] Z. Cao, J. Jerald, and R. Kopper. 2018. Visually-Induced Motion Sickness Reduction via Static and Dynamic Rest Frames. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Reutlingen, Germany, 105–112. https://doi.org/10.1109/VR.2018.8446210

[10] Linchiat Chang and Jon Krosnick. 2010. Comparing Oral Interviewing with Self-Administered Computerized Questionnaires: An Experiment. *Public Opinion Quarterly - PUBLIC OPIN QUART* 74 (03 2010), 154–167. https://doi.org/10.1093/poq/nfp090

[11] HP Development Company. 2020. HP Reverb Virtual Reality Headset - Professional Edition QuickSpecs. https://www.hp.com/h20195/v2/GetDocument.aspx?docname=c06227520

[12] Donald Degraen, André Zenner, and Antonio Krüger. 2019. Enhancing Texture Perception in Virtual Reality Using 3D-Printed Hair Structures. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300479

[13] Darragh Egan, Sean Brennan, John Barrett, Yuansong Qiao, Christian Timmerer, and Niall Murray. 2016. An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, Lisbon, Portugal, 1–6. https://doi.org/10.1109/QoMEX.2016.7498964

[14] A. S. Fernandes and S. K. Feiner. 2016. Combating VR sickness through subtle dynamic field-of-view modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, Greenville, USA, 201–210.

[15] Julian Frommel, Katja Rogers, Julia Brich, Daniel Besserer, Leonard Bradatsch, Isabel Ortinau, Ramona Schabenberger, Valentin Riemer, Claudia Schrader, and Michael Weber. 2015. Integrated Questionnaires: Maintaining Presence in Game Environments for Self-Reported Data Acquisition. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (London, United Kingdom) *(CHI PLAY '15)*. Association for Computing Machinery, New York, NY, USA, 359–368. https://doi.org/10.1145/2793107.2793130

[16] D. Gall and M. E. Latoschik. 2018. The Effect of Haptic Prediction Accuracy on Presence. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Reutlingen, Germany, 73–80.

[17] Ceenu George, Philipp Janssen, David Heuss, and Florian Alt. 2019. Should I Interrupt or Not? Understanding Interruptions in Head-Mounted Display Settings. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 497–510. https://doi.org/10.1145/3322276.3322363

[18] Nathan Navarro Griffin, James Liu, and Eelke Folmer. 2018. Evaluation of Handsbusy vs Handsfree Virtual Locomotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 211–219. https://doi.org/10.1145/3242671.3242707

[19] Nico Haas. 2017. Evaluation of 'In-VR-Questionnaires'. https://doi.org/10.18419/opus-10022

[20] Florian Hruby. 2015. The Sound of Being There: Audiovisual Cartography with Immersive Virtual Environments. *Journal of Cartography and Geographic Information* 69 (2015), 19–28. https://doi.org/10.1007/s42489-019-00003-5

[21] Brent Insko. 2003. 7 Measuring Presence: Subjective, Behavioral and Physiological Methods. *Being there: concepts, effects, and measurement of user presence in synthetic environments* 5 (01 2003).

[22] P. Jordan. 2008. Auswahl einer geeigneten Methode zur Usability Evaluation. http://www2.hawaii.edu/~philippj/resources/Usability_Jordan.pdf

[23] HyeongYeop Kang, Geonsun Lee, Seongsu Kwon, Ohung Kwon, Seongpil Kim, and JungHyun Han. 2018. Flotation Simulation in a Cable-Driven Virtual Environment – A Study with Parasailing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3174206

[24] L. Kruse, E. Langbehn, and F. Steinicke. 2018. I Can See on My Feet While Walking: Sensitivity to Translation Gains with Visible Feet. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Reutlingen, Germany, 305–312.

[25] Steven M. LaValle. 2019. Virtual Reality. http://vr.cs.uiuc.edu/

[26] Benny Liebold, Michael Brill, Daniel Pietschmann, Frank Schwab, and Peter Ohler. 2016. Continuous Measurement of Breaks in Presence: Psychophysiology and Orienting Responses. *Media Psychology* 20 (07 2016), 1–25. Issue 3. https://doi.org/10.1080/15213269.2016.1206829

[27] Ultrahaptics Ltd. 2020. Leap Motion Documentation. https://developer.leapmotion.com/documentation

[28] Philipp Mayring. 2000. Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal], http://qualitative-research.net/fqs/fqs-e/2-00inhalt-e.htm* 1 (06 2000).

[29] Sebastian Oberdörfer, David Heidrich, and Marc Erich Latoschik. 2019. Usability of Gamified Knowledge Learning in VR and Desktop-3D. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300405

[30] Susanne Putze, Dmitry Alexandrovsky, Felix Putze, Sebastian Höffner, Jan David Smeddinck, and Rainer Malaka. 2020. Breaking The Experience: Effects of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376144

[31] Daniel Roth and Marc Erich Latoschik. 2019. Construction of a Validated Virtual Embodiment Questionnaire. arXiv:1911.10176 [cs.HC]

[32] Majed Samad, Elia Gatti, Anne Hermes, Hrvoje Benko, and Cesare Parise. 2019. Pseudo-Haptic Weight: Changing the Perceived Weight of Virtual Objects By Manipulating Control-Display Ratio. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300550

[33] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence* 10 (06 2001), 266–281. https://doi.org/10.1162/105474601300343603

[34] Valentin Schwind, Pascal Knierim, Lewis Chuang, and Niels Henze. 2017. "Where's Pinky?": The Effects of a Reduced Number of Fingers in Virtual Reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Amsterdam, The Netherlands) *(CHI PLAY '17)*. Association for Computing Machinery, New York, NY, USA, 507–515. https://doi.org/10.1145/3116595.3116596

[35] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300590

[36] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1577–1582. https://doi.org/10.1145/3025453.3025602

[37] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50, 6, Article 96 (2017), 39 pages. https://doi.org/10.1145/3134301

[38] Mel Slater. 1999. Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence (Camb.)* 8 (12 1999). https://doi.org/10.1162/105474699566477

[39] Mel Slater and Anthony Steed. 2000. A Virtual Presence Counter. *Presence: Teleoper. Virtual Environ.* 9, 5 (Oct. 2000), 413–434. https://doi.org/10.1162/105474600566925

[40] Unity3d. 2020. User Interfaces for VR. https://unity3d.com/de/learn/tutorials/topics/virtual-reality/user-interfaces-vr

[41] J.D. Van der Laan, A. Heino, and D. De Waard. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research - Part C: Emerging Technologies* 5, 1 (1997), 1–10.

[42] Alexandra Voit, Sven Mayer, Valentin Schwind, and Niels Henze. 2019. Online, VR, AR, Lab, and In-Situ: Comparison of Research Methods to Evaluate Smart Artifacts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300737

[43] C. Wienrich, K. Schindler, N. Döllinqer, S. Kock, and O. Traupe. 2018. Social Presence and Cooperation in Large-Scale Multi-User Virtual Reality - The Relevance of Social Interdependence for Location-Based Environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Reutlingen, Germany, 207–214.

[44] B.G. Witmer and M.J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (1998), 225-240. https://doi.org/10.1162/105474698565686

[45] Bob G. Witmer, Christian J. Jerome, and Michael J. Singer. 2005. The Factor Structure of the Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 14 (2005), 298–312.