



Article

# Assessing Heuristic Evaluation in Immersive Virtual Reality—A Case Study on Future Guidance Systems

Sebastian Stadler <sup>1,\*</sup>, Henriette Cornet <sup>2</sup> and Fritz Frenkler <sup>3</sup>

<sup>1</sup> Faculty of Media, University of Applied Sciences Ansbach, 91522 Ansbach, Germany

<sup>2</sup> UITP—International Association of Public Transport, B-1080 Brussels, Belgium

<sup>3</sup> Senior Excellence Faculty, Technical University of Munich, 80333 Munich, Germany

\* Correspondence: sebastian.stadler@hs-ansbach.de; Tel.: +49-981-4877-235

**Abstract:** A variety of evaluation methods for user interfaces (UI) exist such as usability testing, cognitive walkthrough, and heuristic evaluation. However, UIs such as guidance systems at transit hubs must be evaluated in their intended application field to allow the effective and valid identification of usability flaws. However, what if evaluations are not feasible in real environments, or laboratorial conditions cannot be ensured? Based on adapted heuristics, in the present study, the method of heuristic evaluation is combined with immersive Virtual Reality (VR) for the identification of usability flaws of dynamic guidance systems (DGS) at transit hubs. The study involved usability evaluations of nine DGS concepts using the newly proposed method. The results show that compared to computer-based heuristic evaluations, the use of immersive VR led to the identification of an increased amount of “severe” usability flaws as well as overall usability flaws. Within a qualitative assessment, immersive VR is validated as a suitable tool for conducting heuristic evaluations involving significant advantages such as the creation of realistic experiences in laboratorial conditions. Future work seeks to further prove the suitability of using immersive VR for heuristic evaluations and compare the proposed method to other evaluative methods.

**Keywords:** virtual reality; heuristic evaluation; human-computer interaction; user interfaces; guidance system



**Citation:** Stadler, S.; Cornet, H.; Frenkler, F. Assessing Heuristic Evaluation in Immersive Virtual Reality—A Case Study on Future Guidance Systems. *Multimodal Technol. Interact.* **2023**, *7*, 19. <https://doi.org/10.3390/mti7020019>

Academic Editors: Mark Billingham, Fotis Liarokapis, Lars Erik Holmquist and Mu-Chun Su

Received: 20 December 2022

Revised: 31 January 2023

Accepted: 7 February 2023

Published: 9 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Evaluation methods such as usability testing, cognitive walkthrough, and heuristic evaluation help researchers and developers to evaluate products such as user interfaces (UI) to gain insights about aspects such as usability, user experience, and cognitive load [1–4]. Due to their technical nature and feasibility aspects, the majority of these evaluation methods, especially in the domain of UI design, are conducted in controlled environments (e.g., labs or meeting rooms) [5]. This fact already involves limitations in data collection as lab situations are artificial and thus, miss out on collecting crucial aspects such as environmental impacts on the user and product during its usage [6]. Even though systems such as mobile UIs generally allow contextual evaluations, challenges arise, such as the standardization of data collection (e.g., due to changing situations from participant to participant) and the prevention of noise for the data collection. For non-mobile UIs, for instance information panels and guidance systems, contextual evaluations involve challenges in terms of effort (e.g., finding suitable spaces for valid data collection without impairing local operations), legal issues (e.g., observation of people and workspaces), or feasibility (e.g., due to technical challenges of the products). Nevertheless, in order to guarantee properly functioning products or systems that meet user needs, contextual evaluations, especially in terms of usability, are necessary [6,7].

Virtual Reality (VR) is a technology that allows people to experience simulated environments in an immersive way and with a sense of presence. In this publication, VR

is defined as a computer-generated simulation that can be interacted with, consisting of images, videos, and/or sound that represents an environment that the viewer can experience by using electronic equipment. Since immersive VR allows the creation of immersive experiences of artificial but seemingly real environments while still maintaining laboratorial conditions, we propose a method to conduct heuristic evaluations (based on [2]) in immersive VR. The present study investigates to what extent the method of immersive heuristic evaluation supports experts to evaluate UIs that normally could not be evaluated in their actual application domain (i.e., guidance systems at transit hubs).

The main aim of this study is to investigate the impact of using immersive VR for identifying usability flaws compared to display-based simulations in consideration of severity of identified usability flaws. Additionally, we aim to investigate the impact of using VR on the method of heuristic evaluation itself. The significance of the present research lies in the combination of contextual heuristic evaluations in laboratorial environments due to the usage of immersive VR and the potential the proposed method can have on future evaluative studies. The underlying hypotheses are:

1. The heuristic evaluation in immersive VR leads to the identification of overall more usability flaws than conventional heuristic evaluations based on two-dimensional simulations.
2. The heuristic evaluation in immersive VR leads to the identification of more usability flaws with high severity than conventional heuristic evaluations based on two-dimensional simulations.
3. The heuristic evaluation in immersive VR allows contextual evaluations of products in an immersive way while still maintaining laboratorial conditions.

After assessing the advantages and limitations of the proposed method, a set of general guidelines for using immersive heuristic evaluation is presented to foster further research and its active usage as a tangible output for the research community. The principal conclusions of the present studies are:

1. The usage of immersive VR for the method of heuristic evaluation led to the identification of overall more usability flaws than conventional heuristic evaluations based on computer-based simulations.
2. More “minor” usability flaws as well as more “severe” usability flaws were identified by the heuristic evaluation in immersive VR than the desktop-based heuristic evaluation.
3. Besides technical limitations, the usage of immersive VR for heuristic evaluation involves a set of significant advantages over the conventional method, such as immersive and real experiences, natural interaction, laboratorial conditions, and the inclusion of filters.

## 2. Related Work

### 2.1. Heuristic Evaluation

Heuristic evaluation is a method used to evaluate the usability of UIs by having experts examine the UI and identify usability flaws based on a predefined set of rules of thumb, called the “heuristics” [2]. This method was firstly described by Nielsen and Molich in 1990 and has since been widely used in the field of human-computer interaction. The general procedure of heuristic evaluation is as follows (based on [2,3]):

1. Definition of a set of heuristics to evaluate the UIs (e.g., heuristics by Nielsen and Molich [2]).
2. Recruitment of a small team of three to five experts in the field of human-computer interaction and UI design. These experts should be experienced in conducting heuristic evaluations and be familiar with the underlying set of heuristics. Optionally, people with expertise of the intended application domain can be involved.
3. Conduct of the evaluation by providing a set of clear tasks and scenarios to complete while interacting with the UI. During the interaction, the experts document findings, observations, ideas, and identified usability flaws in consideration to the heuristics.

4. Group discussion and consolidation of individual findings in order to derive an agreement on the identified usability flaws including their occurrence and severity (i.e., consensus meeting).
5. Compilation of a report that summarizes the findings of the evaluation.

The heuristics defined by Nielsen and Molich [2] have been adapted and modified for specific case studies by other researchers [5,8]. Mankoff et al. [5] used a modified set of heuristics to evaluate the usability of ambient displays and subsequently compared the modified heuristics with the heuristics that were proposed by Nielsen and Molich [2]. The results show that the tests using the modified heuristics led to the identification of overall more problems, including more severe usability flaws. This fact indicates the advantages of adapting the set of heuristics for specific case studies and technologies.

Heuristic evaluation is considered to be an effective method regarding the identification of usability flaws in a wide range of HCI related topics, including websites [8] and mobile apps [9]. The key advantages of the methods lie in its efficiency in terms of time and cost, as well as the identification of usability flaws that are considered “severe” [2,3,10,11].

## 2.2. Evaluation Methods in Immersive Virtual Reality

Immersive VR has already been utilized for conducting evaluative studies such as usability testing, user experience evaluation, comparative analyses, and experience simulations, as the creation of immersive and realistic environments has the potential to mimic real-world scenarios [9–14]. As immersive VR technology becomes more prevalent, it is increasingly being used as a tool for conducting evaluative studies, including methods such as usability testing, user experience (UX) testing, and observational studies. Case studies include pedestrian safety research, evaluation of UIs inside vehicles, and research in architecture.

Carlsson and Sonesson conducted a UX study in immersive VR in which participants (i.e., potential users and lead users) experienced cockpit UIs inside a vehicle [15]. The results show that immersive VR offered great opportunities for creating engaging and immersive experiences. The researchers defined a set of guidelines for the development of an immersive VR simulator for UX including aspects such as the usage of suitable narratives, ensure sensory congruence, and enabling a range of concepts to be tested. Stadler et al. conducted usability tests for evaluating explicit human-machine interfaces in pedestrian research [16]. The researchers conclude that immersive VR is advantageous for usability testing in terms of safety for participants, validity of results, feasibility of scenarios as well as efficiency in terms of time and costs. The research involved limitations mostly of technical nature such as restricted field of view and display resolution of the head-mounted displays (HMD). Mäkelä et al. conducted two empirical user studies in which the behaviors of participants was compared when seeing public displays once in a physical environment and once in immersive VR [17]. The researchers observed high congruence in behaviors and even increased motivation to explore the environment in immersive VR.

In conclusion, immersive VR is being increasingly used as a tool for conducting evaluative studies across a wide range of fields. The immersive nature of VR allows the creation of seemingly realistic environments and scenarios making it a valuable tool for studying decision-making and human behavior. Overall, studies in which VR was utilized for evaluative purposes suggest that immersive VR turned out to be a valid, cost-effective, and flexible alternative to conventional methods.

## 2.3. Heuristic Evaluation in Virtual Reality

The method of heuristic evaluation has already been applied in the context of immersive VR [8,18–21]. Sutcliffe and Gault evaluated UIs of immersive VR applications based on a set of adapted heuristics [8]. Similarly, Gabbard et al. proposed an iterative human-centered design approach including heuristic evaluation of immersive virtual environments [18]. Silvennoinen and Kuparinen [21] conducted a usability study (i.e., heuristic usability analysis) of medical simulators in immersive VR. The researchers conclude that

the simulator could constitute a promising learning and training tool when usability issues are improved, concluding that user-friendliness is a key aspect for the simulator's effectiveness. In a more recent study, researchers investigated the performance of conducting heuristic evaluations in immersive VR [22]. In this context, a product (i.e., a health box) was evaluated by nine participants both as a physical product as well as a digital product in immersive VR. Even though the researchers did not find statistical differences between the two evaluation methods, insights were gathered regarding the usage of the immersive VR technology. The hardware limitations (i.e., resolution and field of view) resulted in perceived usability flaws that were mentioned by the participants as false positives. Therefore, the researchers suggest that future work should focus on the impact of the representation fidelity on the respective method.

Further studies focused on using heuristics for the evaluation of immersive VR systems. Rusu et al. developed a set of overall 16 usability heuristics as well as a usability checklist for virtual worlds, arguing that these virtual worlds need new evaluation methods [23]. Nevertheless, the researchers do not specifically apply the method of heuristic evaluation in this context. In another study, Billow and Cottam [24] investigated the heuristics by Nielsen and Molich [2] in the context of immersive analytic systems in VR. By applying the method of heuristic evaluation for evaluating an immersive data visualization tool, the researchers conclude that the majority of heuristics still apply for the chosen context whereas several occurring issues revolved around areas where the heuristics did not provide specific guidance. Thus, the researchers suggest that the heuristics need to be modified in the chosen application domain.

In summary, researchers who investigated the usage of heuristics and the method of heuristic evaluation in the context of immersive VR conclude that further research is needed in order to generate general conclusions in this context [18,23,25].

The method of heuristic evaluations has also been investigated in the context of Augmented Reality (AR) and Mixed Reality (MR). Labrie and Cheng investigated the applicability of general heuristics (i.e., Nielsen and Molich [2]) for AR in the context of home design apps. The researchers conclude that the technical limitations of AR can impact the usability of AR applications and since AR is an emerging technology proper onboarding processes for users are necessary [26]. Sari et al. used the method of heuristic evaluation to evaluate the usability of a mobile AR app for designing houses [27]. The method allowed the researchers to identify usability shortcomings of the tested application, especially regarding the heuristics "consistency and standards" as well as "error prevention". Nevertheless, since the present research explicitly focusses on the usage of immersive VR for heuristic evaluation, AR and MR will not further be considered.

### 3. Materials and Methods

In the present chapter, firstly the underlying case study for the experiments is shared, followed by a detailed description of the methodological approach of the present investigation.

#### 3.1. Case Study: Guidance Systems at Bus Interchanges

The chosen case study considered dynamic guidance systems (DGS) at transit hubs in an era of level 5 autonomous mobility. It is anticipated that due to advantages connected to automation, such as bus platooning and optimized schedules, the localization of buses towards specific berths at a bus transit hub may vary throughout the day (similar to an airport where several airlines share the same gate throughout the day). Therefore, passengers at transit hubs will not be able to rely on prior spatial knowledge to board buses anymore but must constantly re-localize the required bus berths. Since this could lead to disorientation, additional cognitive effort and decreased user acceptance, improved DGS concepts were tested to evaluate their usability for leading people to the respective destination berth at a transit hub.

### 3.2. Adaptation of Heuristics

The set of general heuristics of Nielsen and Molich [2] was used as a basis for the present evaluation. It consists of ten heuristics including aspects such as providing a simple and natural dialogue, consistency, and feedback. Researchers have already used the general heuristics of Nielsen and Molich in an adapted way to ensure suitability for the chosen case study [5,8]. Since heuristics such as “Provide shortcuts” do not apply to the present study, a workshop was conducted in which the chosen experts for the experiments had the possibility to adapt the heuristics to be tailored upon the case study’s needs. During this workshop, the experts experienced the virtual environment including a baseline guidance system (i.e., the current UI of the transit hub) and were asked to adapt the heuristics accordingly. The following heuristics were defined for the chosen case study:

- Continuity of information (i.e., the DGS should provide information at the transit hub as continuously as possible. Thus, regardless of the user’s location at the transit hub, guidance information should be provided.)
- Consistency of information (i.e., the DGS should be consistent in terms of provided information including its visual representation.)
- Visibility of information (i.e., the placement and size of DGS provides sufficient visibility for users at the transit hub.)
- Adequate information (i.e., the DGS provides adequate and necessary information to ensure successful guidance inside the transit hub.)
- Comprehensibility of information (i.e., the information displayed by the DGS is understandable to the user.)
- Intuitive and clear interface (i.e., the DGS is designed to foster usability and reduce cognitive load.)
- Accessibility for vulnerable users (i.e., the DGS should be usable for vulnerable users—in this study this refers to seniors.)
- Provision of feedback (i.e., the DGS implicitly or explicitly provides feedback to users.)
- Prevention of errors (i.e., the DGS is designed in a way to reduce the risk of wrong interaction for users.)

Overall, two heuristic evaluations were conducted: one in immersive VR and one with the help of a computer-based simulation that was visualized on a screen. This allowed a direct assessment of the impact that the technology of immersive VR had on the method. Both methods are described in detail in the following subsections.

### 3.3. Development of the VR Application

Prior to the conduct of the heuristic evaluation, an immersive VR application was developed to represent the application field of the test and offer adequate interaction techniques for the experts to experience the proposed DGS. The application was built using the game engine Unity 3D 2019.1.3. The development involved the expertise of one designer, two software developers, as well as one 3D visualizer and took four months of development and optimization time until its completion. To ensure a realistic environment, a digital twin of a transit hub in Singapore was created (i.e., Boon Lay Station). The necessary information regarding the transit hub was collected by on-site visits and documentation. Figure 1 shows the virtual environment including a crowd simulation and baseline guidance system (i.e., the existing guidance system at the transit hub).

Necessary interaction techniques were implemented for allowing effective and efficient evaluations of the proposed UIs by the experts. This involved the inclusion of locomotion methods within a six degrees-of-freedom setup (i.e., touchpad movement, teleportation, arm-swing movement, and physical movement) and possibilities to adapt the environment (e.g., changing the crowd size at the transit hub and adapt arrival time of busses). This functionality was implemented to be accessible via a compressible floating UI. Figure 2 shows two screens of the immersive VR UI for modifying the scene and environment (i.e., left: the interface for changing aspects such as crowd level and locomotion technique; right: the interactive map with the option to teleport to the entrances of the transit hub). A precise

and effective interaction was ensured due to ray-casts and by using the input devices of the respective HMD.



Figure 1. Digital twin of a Singaporean transit hub including a baseline guidance system.

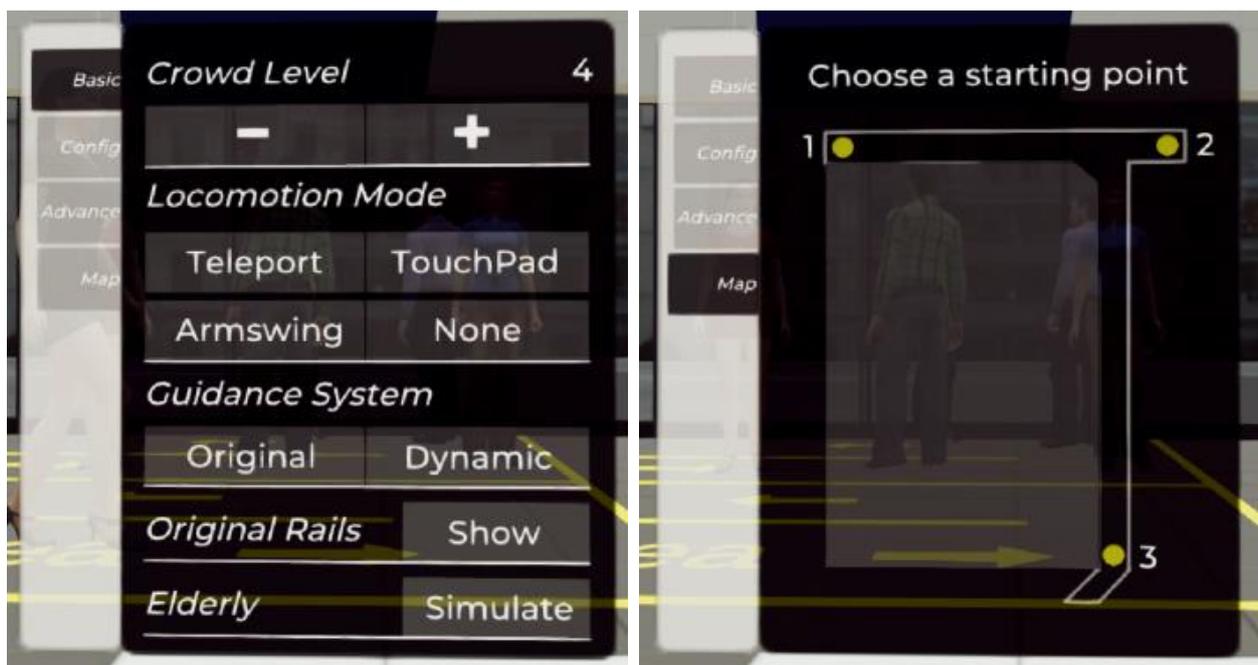


Figure 2. User interfaces of the immersive VR simulation to change and adapt the environment.

The following functionalities were included in the immersive VR simulation:

- Nine DGS concepts were implemented and accessible via the immersive VR UI.
- A crowd simulation including three-dimensional humanoid agents was integrated, representing crowd movements throughout the day. The crowd density could manually be increased and decreased by the experts. The crowd simulation was realized by using NavMesh.
- Five scenarios were implemented involving the completion of specific tasks in the virtual environment to allow experts to experience the DGS concepts in specific situations (see Section 3.8).

A further focus of the application was the integration of a mode that simulates the environment out of the perspective of vulnerable users (in this context elderly people).

This was realized by including a mode that simulates an occlusion of peripheral vision as well as reducing the walking speed of touchpad movement and arm swing movement to 60%. Therefore, the experts were able to experience the virtual environment and proposed guidance systems out of the perspective of vulnerable users to investigate the system's accessibility.

### 3.4. Study Setup

The experiments were conducted in a controlled environment inside a conference room with an empty tracked area of 6.2 m by 3.8 m to allow six degrees-of-freedom tracking for the HMD. An HTC Vive Pro HMD was used with HTC Controller and HTC Lighthouse 2.0 trackers. A high-performance notebook was used as hardware for running the simulation (i.e., HP Omen i7-8750H, NVIDIA GeForce GTX 1070, 32 GB DDR4-2666 SDRAM, 512 GB PCIe nvmetm M.2 SSD).

### 3.5. Recruitment of Experts

The following structured approach was chosen to identify and recruit experts for the heuristic evaluation:

1. The necessary expertise and required experience for the experts were defined including the field of UI, UX and usability [2,3]. Beyond that, experts from the intended application field for the guidance systems were required as evaluations in this application domain could lead to specific challenges that require deep understanding of the operation of the transit hub. To ensure a high degree of expertise and experience, at least five years of working experience in the required context was presupposed. Additionally, experts from industries, academia, and government were preferred for allowing different perspectives on the DGS. As the workshops were held in Singapore, experts from Singapore were especially searched. Since research suggests that a small sample size of five to ten participants is already sufficient to identify the majority of usability flaws, overall, six experts were identified and recruited for the heuristic evaluations (based on [2,28]).
2. By conducting online research, accessing the established professional network of the researchers as well as actively approaching governmental institutes and universities, overall, six experts were identified. This included two experts in the field of design, usability, and UI from industry, two experts from the field of architecture with knowledge in UI and UX from academia, as well as two experts from local traffic authorities in Singapore (government).
3. After approaching the identified experts, screening questionnaires were handed to ensure suitability of expertise and experience for the planned workshops.
4. As all identified experts passed the screening questionnaire, these six experts were invited to the heuristic evaluation workshops (16.7% female, 83.3% male; age: 24–64; M = 36.17; S.D. = 14.39).

Since the designers and architects had expertise in UI, UX, and usability, they were paired with experts in public transport to ensure multidisciplinary groups. All experts had moderate experience in immersive VR. The experts were grouped into pairs of two with one UI expert included. Consequently, for the conduct of the heuristic evaluation in immersive VR, there were three groups consisting of two experts each, in which at least one was trained in UI, UX, and usability.

The recruitment of experts for the comparative heuristic evaluation (by utilizing the computer-based simulation) corresponded with the above-mentioned procedure. Thus, six experts from the field of Design, UI, UX, and usability, as well as architects and public transport experts were involved in the study (16.7% female, 83.3% male; age: 26–59; M = 34; S.D. = 14.26).

### 3.6. Study Procedure

After a formal introduction to the case study, the method, the interfaces, and immersive VR (~0.5 h), the experts had time to familiarize themselves with the technology of immersive VR (~0.25 h). After clarifying questions, the experts were paired and started the heuristic evaluation. While one expert experienced the virtual environment, the other expert followed the events on a screen and took notes from the statements of the expert in immersive VR who followed a “think aloud protocol” (i.e., spontaneously sharing all observations, findings, usability flaws, and ideas while experiencing the application; based on [29]). Figure 3 shows one expert undergoing a scenario and sharing insights while the other expert observes the video footage (showing a first-person view) and notes down statements from the expert immersed in VR.



**Figure 3.** Two experts during a scenario of the heuristic evaluation.

After one expert experienced all UIs in all the scenarios, the experts swapped roles and followed the same procedure again. Once both experts experienced all UIs in all scenarios, the interactive part of the heuristic evaluation was finished (~1 h). Subsequently, the experts had time to discuss their observations within their group before getting back together with the other groups to discuss the experiences among all experts (~1.5 h). The main objective of the expert discussion was to share usability flaws of the UIs, ideas, and observations that were identified during the interactive part in immersive VR. The experts prepared a report to document the concluded findings of the heuristic evaluation as precisely as possible. Finally, the experts shared their opinions regarding the usage of immersive VR for heuristic evaluation (~0.5 h).

### 3.7. DGS Concepts

Overall, nine DGS concepts were included in the tests. The intended application field for the concept was the transit hub that functioned as basis for the experiments (i.e., Boon Lay Station in Singapore). To ensure a variety of concepts, a range of technologies was used for the concepts, including screens, information points, LED stripes, and AR applications on Smartphones. Figure 4 shows a summary of the tested DGS concepts.

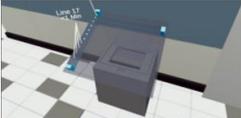
Concept	Description	Picture	Concept	Description	Picture
1	Screens with icons, colours, and patterns including secondary information		6	Information points that show information via a projection	
2	One screen per bus line with secondary information		7	LED stripes that are integrated in the floor (one coloured stripe per bus line)	
3	AR app that shows the destination berth even through walls		8	Personalized information on the floor in front of the passenger via LED tiles	
4	AR app that shows a rolling ball that guides the user to the destination berth		9	A personal assistance robot that leads the passenger to the destination berth	
5	AR app in addition to overhead signage that excludes unnecessary information				

Figure 4. Summary of DGS concepts tested via the heuristic evaluation in immersive VR.

### 3.8. Scenarios

Scenarios were implemented to allow the experts to experience each DGS in specific situations. Each scenario included a pre-defined starting point as well as a task to complete. The tasks to complete were as follows:

- Start at starting point two and try to enter the bus that drives to Boon Lay Drive
- Start at starting point one and board a bus that drives to Bukith Merah
- Start at destination berth B13 and transfer to the bus that drives to Bukith Merah
- Start at starting point three, collect a book at berth B09 and subsequently board a bus that drives to Boon Lay Drive
- Start at starting point three and find the bus that drives to Jurong East

The arrival times of all buses differed among the scenarios in order to increase the variability for investigating each UI.

### 3.9. Computer-Based Heuristic Evaluation

The computer-based heuristic evaluation followed the same procedure as the heuristic evaluation in immersive VR to ensure direct comparability. However, the experts evaluated the guidance systems alone and documented identified usability flaws in a virtual notebook that was implemented in the computer-based application. The tests were conducted in a controlled environment in a conference room. The same computer hardware was used for the computer-based simulation as for the heuristic evaluation in immersive VR. A keyboard- and mouse-based locomotion was chosen for moving in the computer-simulated environment. After the tests, a group discussion was conducted to conclude and derive a study report.

## 4. Results

Due to the small sample size of the heuristic evaluation, a descriptive analysis of results was chosen to directly compare the identified usability flaws both in immersive VR and the desktop-based application.

### 4.1. Usability Flaws Identified in Immersive VR

Table 1 shows a summary of usability flaws that were documented by the experts during the heuristic evaluation and subsequent group discussion. The identified usability

flaws are clustered by the adapted set of heuristics as well as the total amount of discovered usability flaws of all nine DGS concepts identified in immersive VR. The identified usability flaws were separated into the two groups of “minor” (M) and “severe” (S) [25].

**Table 1.** Overview of identified usability flaws discovered among the nine guidance systems in immersive Virtual Reality.

Heuristic	DGS Concepts																		Total per Concept	
	1		2		3		4		5		6		7		8		9		M	S
	M	S	M	S	M	S	M	S	M	S	M	S	M	S	M	S				
Continuity of information		1		1		1				1		1					1		1	5
Consistency of information	1						1		1		1								4	0
Visibility of information	1	1			1					1	1	1	1		1	1			5	4
Comprehensibility of information	3	2											1		1				5	2
Intuitive and clear interface		1	1										1		1				3	1
Adequate information	1	1	1		1				1			1	1			1			5	3
Accessibility for vulnerable users	1				1		1		1		1				1			1	6	1
Provision of feedback	1		1		1		1		1		1				1			1	9	0
Prevention of errors	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8	3
<b>Total number of flaws</b>	<b>9</b>	<b>6</b>	<b>4</b>	<b>1</b>	<b>5</b>	<b>2</b>	<b>4</b>	<b>1</b>	<b>5</b>	<b>3</b>	<b>5</b>	<b>3</b>	<b>6</b>	<b>0</b>	<b>6</b>	<b>2</b>	<b>3</b>	<b>1</b>	<b>46</b>	<b>19</b>

M = Minor; S = Severe.

A total of 19 out of the overall 65 identified usability flaws were retrospectively clustered as “severe” usability flaws by the experts and researchers while 46 usability flaws were defined as “minor” flaws. A usability flaw was considered “severe” if the flaw actively influenced decision making due to, for instance, wrong or unclear information or due to the absence of information in critical moments.

4.2. Usability Flaws Identified via the Computer Simulation

Table 2 shows a summary of usability flaws, clustered by the adapted set of heuristics as well as the total amount of discovered usability flaws of all nine DGS concepts identified via the computer simulation. The identified usability flaws were separated into the two groups of “minor” (M) and “severe” (S) [25].

**Table 2.** Overview of identified usability flaws discovered among the nine guidance systems via the computer simulation.

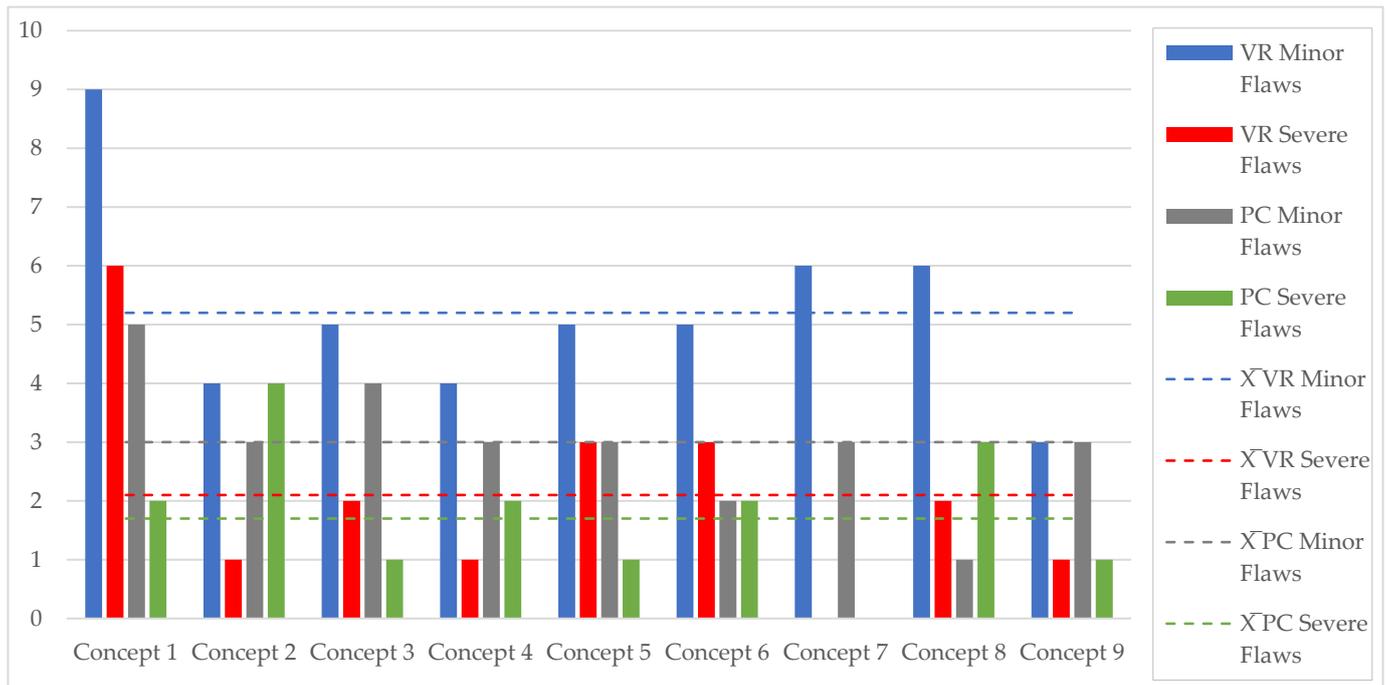
Heuristic	DGS Concepts																		Total per Concept	
	1		2		3		4		5		6		7		8		9		M	S
	M	S	M	S	M	S	M	S	M	S	M	S	M	S	M	S				
Continuity of information	1	1		1	1		1								1				3	3
Consistency of information	1			1									1				1		3	1
Visibility of information	1		1	2		1			1			1			1				3	5
Comprehensibility of information		1							1		1						1		3	1
Intuitive and clear interface	1		1				1		1										4	0
Adequate information					1							1			1				1	2
Accessibility for vulnerable users							1						1					1	2	1
Provision of feedback	1				1		1				1		1				1		5	1
Prevention of errors			1		1		1		1		1				1				3	2
<b>Total number of flaws</b>	<b>5</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>4</b>	<b>1</b>	<b>3</b>	<b>2</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>3</b>	<b>0</b>	<b>1</b>	<b>3</b>	<b>3</b>	<b>1</b>	<b>27</b>	<b>16</b>

M = Minor; S = Severe.

A total of 16 out of the overall 43 identified usability flaws were retrospectively clustered as “severe” usability flaws by the experts and organizations while 27 usability flaws were defined as “minor” flaws. The procedure of defining a usability flaw as “minor” or “severe” followed the same criteria as for the heuristic evaluation in immersive VR.

#### 4.3. Comparison of Identified Usability Flaws Based on Technology

Figure 5 shows a graph of the overall identified usability flaws per DGS concept, divided into “minor” and “severe” usability flaws. Furthermore, the mean values of identified flaws among all concepts are displayed as dotted lines.



**Figure 5.** Comparison of identified usability flaws and its means based on display technology (i.e., desktop-based vs. immersive VR).

Figure 5 shows that the overall amount of identified usability flaws including the identification of “minor” and “severe” usability flaws was higher while using immersive VR than using the display-based computer simulation. Nevertheless, even though due to the usage of immersive VR overall more minor and severe usability flaws were identified, proportionally more severe usability flaws with high severity were identified by using the desktop-based method (37.21%) compared to immersive VR (29.23%).

#### 4.4. Qualitative Assessment of the Heuristic Evaluation in Immersive VR

Subsequent to the experiments, a group discussion was conducted in order to assess the usage of immersive VR for heuristic evaluations (i.e., consensus meeting). The following advantages of using the technology for this method were defined by the experts:

- Immersive VR allows a very realistic and immersive experience.
- Scenarios and concepts can be created and experienced in a way that would not be possible in real-life conditions.
- Immersive VR allows standardized evaluations of concepts in its intended application domain(s).
- Immersive VR makes a big difference to these kinds of workshops due to its fidelity of spatial awareness and scale.
- Without immersive VR, the experts would need more time to understand the case study and to experience all concepts.

- Immersive VR allows a fast immersion into the environment without the need for on-site visits.
- Immersive VR is a good tool to experience environments and concepts.
- One key advantage is the usage of “elderly mode”, which is exclusive to immersive VR. People can experience an approximation of how it would feel to be an elderly user (e.g., with visual occlusion and reduced walking speed). Immersive VR offers the only way to experience that.
- The range of locomotion techniques is very advantageous; with the teleportation locomotion, the environment can be investigated very quickly, whereas with the physical walking method and arm swing method, the effort, distance, and time to reach certain points in VR can be experienced very realistically.
- The immersive VR devices, including its input devices and the VR interface, are easy to use.

Furthermore, the following challenges and limitations of using VR for the specific heuristic evaluations were identified and mentioned by the experts:

- Participants with little VR experience should not stay too long in VR to minimize the risk of cybersickness.
- The expert who takes notes can get nauseous by watching the video footage of the first-person view (due to an unsteady field of view in immersive VR).
- The movement sensitivity with the touchpad locomotion technique is too high and therefore, using the touchpad for moving was not as realistic as the other locomotion techniques.
- Continuous rotation while using touchpad movement fosters cybersickness. Thus, it should be changed into incremental turning (i.e., changing the perspective by a pre-determined angle).
- When the crowd level is set to “high”, the performance (i.e., framerate) decreases, which could lead to cybersickness.
- In other study contexts, the absence of haptic feedback could cause limitations while evaluating concepts.
- Technical limitations of the HMD decrease evaluation capabilities (e.g., visibility of information due to restricted field of view and resolution).

#### 4.5. Design Guidelines for Conducting Heuristic Evaluations in Immersive VR

Based on the observations during the heuristic evaluation in immersive VR as well as the group discussion, the following guidelines were defined for fostering the development and conduct of heuristic evaluations in immersive VR:

- Ensure the involvement of researchers with sufficient expertise regarding the hardware, software and study context since participants might need assistance.
- Take precautions to comfort participants in case of cybersickness.
- Ensure a controlled environment and laboratorial conditions for valid and reliable data collections.
- Include a wide range of concepts for the tests: Since immersive VR allows the evaluation of different concepts and variants without additional effort, its full potential should be used to identify a range of usable concepts and variants.
- Include a range of conditions/scenarios and specific tasks: Since immersive VR allows to change the environment and conditions very quickly and easily, the concepts and variants should be tested in different conditions to increase the evaluation potential.
- Prioritize content interactivity over visual representation: For participants, a precise and natural way of interacting with the system is more important than its visual representation for evaluating concepts (i.e., willing suspension of disbelief).
- Avoid the integration of aspects that would expose technical limitations of immersive VR (e.g., the resolution of the displays of the used HMDs and their field of view).

## 5. Discussion

In the present study, a method for conducting heuristic evaluations in immersive VR is proposed. The contribution of the current investigation is twofold: Firstly, results are shared that indicate that the usage of immersive VR for the conduct of heuristic evaluations has the potential to allow experts to identify overall more usability flaws compared to computer-based simulations. In addition, not only more “minor” usability flaws but also more “severe” usability flaws could be identified thanks to the immersive technology. Nevertheless, proportionally more severe usability flaws were identified through the desktop-based heuristic evaluation compared to the immersive VR method. Possible explanations for this fact could be that the immersive experience of the environment fosters the identification of additional minor flaws that do not become apparent in the desktop-based application or that are even caused by the immersive technology (e.g., due to its resolution and field of view). This fact could indicate the potential risk of identifying pseudo-problems by using immersive VR. This is in line with Zhang and Simeone who concluded that the usage of immersive VR for heuristic evaluation and its technical limitations could lead to false positives (i.e., the identification of invalid usability flaws) [22]. However, the validity of this fact needs to be investigated in future research iterations. Secondly, by a qualitative assessment of the method by the involved experts as well as the research team, a set of guidelines was derived that functions as a basis for fostering further research as well as the development of VR heuristic evaluations. Beyond the potential for enhancing the identification of usability flaws, the assessment showed great advantages of using VR, such as the possibility to create immersive and realistic scenarios for investigating concepts in certain application domains in laboratorial conditions, the reproducibility of scenarios for every expert, the possibility to adapt crowd levels on demand, as well as the integration of modes that let the experts experience each concept out of the perspective of vulnerable users. In addition to the above stated advantages, the advantages of conventional heuristic evaluations, such as time- and cost-efficiency, were still maintained (based on [3]), since the majority of tasks for developing and preparing the workshops only insignificantly differ from the conventional method.

In a method comparison study, Jeffries et al. [3] found that heuristic evaluation is an inexpensive evaluation method that offers efficiency in terms of identification of usability flaws. Furthermore, Nielsen and Molich [2] stated that the major advantages of heuristic evaluation are cost-effectiveness, the intuitive procedure, no need for advanced planning, and the fact that the method can be used in the early phases of the development process. This is consistent with the present findings, since with immersive VR the conduct of a heuristic evaluation in an inexpensive and effective way without the need for an on-site evaluation or physical prototyping was possible. Several research projects already investigated the suitability of using VR for concept evaluations. Researchers investigated for instance pedestrian safety aspects and explicitly human-machine interfaces as communication cues for crossing scenarios at zebra crossings in an era of fully autonomous vehicles [16,30–32]. The researchers conclude that immersive VR experiences offer great advantages such as feasibility of scenarios that would not be replicable in real-life conditions, the collection of valid and reliable data, as well as time- and cost-efficiency. These findings are consistent with the advantages that became evident in the present investigation. Further research suggests that immersive VR can constitute a suitable tool for evaluative studies such as usability tests, behavioural observations, and user experience studies. Kuliga et al. [33] compared user experience aspects in a real building with its digital twin in immersive VR. The researchers concluded that there were no significant differences in UX ratings between the real-building and its digital representation indicating that in this context VR could be a suitable empirical research tool. These findings are in line not only with the quantified results of the present study that indicate that immersive VR can improve evaluative capabilities but also the qualitative assessment in which experts expressed the suitability and advantages of using immersive VR for this method.

The conduct of the present method also involved a set of limitations. Firstly, the experts claimed that spending too much time in immersive VR could lead to cybersickness. Secondly, the experts mentioned that by setting the crowd level to “high”, the performance of the immersive VR application decreased. Since the framerate of the immersive VR application is directly affected by this aspect, the risk of cybersickness is increased [34]. Additionally, due to the small test population for the present comparative analysis, a generalization of results is not possible. Thus, the present results are exploratory and should be treated as suggestions. Further studies are required, to prove the suitability of using immersive VR for heuristic evaluations in a valid and reliable way. In this context, as only one application domain was chosen for the present study, impacts of using VR for such methods in other application domains need to be proven in future studies as well.

The implications of the present study are as follows. Especially when heuristic evaluations are not feasible in real-life conditions or would involve a considerable amount of effort, time, and money, the usage of immersive VR could constitute an advantageous alternative for evaluating concepts in terms of usability in its intended application domains. Especially compared to non-immersive simulations, immersive VR offers great advantages, as it allows experts to experience concepts in a realistic and immersive way. Moreover, in contrast to contextual evaluations, the proposed method can be conducted in a laboratorial environment that ensures controlled conditions for data collection. Overall, the results indicate that immersive VR has the potential to become a valid research tool for evaluative methods such as heuristic evaluations.

Future research is planned to further quantify the impact of immersive VR on heuristic evaluation but also to determine its effectiveness and efficiency compared to other evaluation techniques such as usability testing and cognitive walkthrough. Furthermore, additional application domains are planned to be included in next iterations for allowing a certain level of generalizability for the proposed method. By conducting further research, it is also planned to revise the first set of design guidelines for the development of heuristic evaluations in immersive VR.

## 6. Conclusions

The evaluation of UIs is a fundamental part of the product development and design process to ensure usability and user acceptance of products. Nevertheless, in order to guarantee properly functioning products that meet user needs, especially contextual evaluations are necessary as products are directly impacted by its surroundings. Considering this, conventional evaluation methods such as usability testing and heuristic evaluation may fall short since these methods are mainly conducted in artificial and isolated environments (especially for UIs with certain dimensions such as information panels or guidance systems). Therefore, in the present study, a method is proposed that allows the conduct of heuristic evaluations in immersive VR. Thus, UIs can be evaluated in their intended application field without the need for on-site experiments. The key findings of the present research are:

1. Based on the expert statements, a basic feasibility of the proposed method in immersive VR for the chosen case study was proven without losing the advantages of the conventional method of heuristic evaluation (e.g., low cost- and time-involvement).
2. Additional advantages were identified due to immersive VR such as a high level of immersion, the consideration of the UIs' application field, controlled conditions in a laboratorial environment, reproducibility and standardization of data collection, and the integration of modes to enable experts to experience the UIs as vulnerable users.
3. More “minor” and “severe” usability flaws were identified by using immersive VR compared to the desktop-based application.

Potential cybersickness caused by immersive VR as well as the limited generalizability of results caused by the small sample size and specific application domain constitute the main limitations of the present study. Overall, the present study implies that immersive VR could be a valuable tool for the conduct of heuristic evaluations especially when physical prototypes are not feasible or available. However, besides technical limitations of VR that

should not be exposed by the application, the potential identification of pseudo-problems by the experts needs carefully to be considered and avoided in order to ensure effectiveness of the heuristic evaluation. Besides the verification of the stated research questions, the main contribution of the present investigation is the proposition of a preliminary set of guidelines for developing and conducting heuristic evaluations in immersive VR. The present study functions as a basis for further research in the field of UI evaluation and human-computer interaction. Future research will focus on further proving the suitability of using immersive VR for heuristic evaluations, test the method in different application domains and compare it to other evaluative methods such as usability testing and cognitive walkthrough.

**Author Contributions:** Conceptualization, S.S. and H.C.; methodology, S.S., H.C. and F.F.; software, S.S.; validation, S.S. and H.C.; formal analysis, S.S.; investigation, S.S. and H.C.; resources, S.S., H.C. and F.F.; data curation, S.S.; writing—original draft preparation, S.S., H.C. and F.F.; writing—review and editing, S.S., H.C. and F.F.; visualization, S.S.; supervision, H.C. and F.F.; project administration, S.S., H.C. and F.F.; funding acquisition, S.S., H.C. and F.F.. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was part of the project “From Virtual Reality to Simulation: User-Centred Design of Dynamic Guidance Systems for Transit Hubs” in collaboration with the Singapore-ETH Centre and Nanyang Technological University Singapore (NTU) funded by the National Research Foundation, Singapore, as part of the Intra-CREATE Seed Collaboration Grant (Project ID: NRF2018-ITS003-015).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors thank Panagiotis Mavros, Rohit Kumar Dubey, Philipp Andelfinger, David Eckhoff, and Xu Hong for the collaboration. The authors further thank all experts for participating in the workshops. Lastly, the authors thank the reviewers and editors of Multimodal Technologies and Interactions for their time and valuable feedback.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lewis, C.; Polson, P.G.; Wharton, C.; Rieman, J. Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 1990; pp. 235–242. [\[CrossRef\]](#)
2. Nielsen, J.; Molich, R. Heuristic Evaluation of User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 1990; pp. 249–256. [\[CrossRef\]](#)
3. Jeffries, R.; Miller, J.R.; Wharton, C.; Uyeda, K. User interface evaluation in the real world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 1991; Volume 91, pp. 119–124. [\[CrossRef\]](#)
4. Rubin, J.; Chisnell, D. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*; Wiley Publishing Inc.: Indianapolis, IN, USA, 2008.
5. Mankoff, J.; Dey, A.K.; Hsieh, G.; Kientz, J.; Lederer, S.; Ames, M. Heuristic evaluation of ambient displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2003; Volume 5, pp. 169–176.
6. Duda, S.; Warburton, C.; Black, N. Contextual Research: Why We Need to Research in Context to Deliver Great Products. *Lect. Notes Comput. Sci.* **2020**, *12181*, 33–49. [\[CrossRef\]](#)
7. Heufler, G. *Design Basics—From Ideas to Products*; Niggli Verlag: Zürich, Switzerland, 2004.
8. Sutcliffe, G.; Gault, B. Heuristic evaluation of virtual reality applications. *Interact. Comput.* **2004**, *16*, 831–849. [\[CrossRef\]](#)
9. Lazar, J.; Feng, J.H.; Hochheiser, H. *Research Methods in Human-Computer Interaction*; Elsevier Morgan Kaufmann: Cambridge, MA, USA, 2017.
10. Nielsen, J. Usability Inspection Methods. In *CHI'94: Conference Companion on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 1994; pp. 413–414. [\[CrossRef\]](#)
11. Guttman, L.A. New Approach to Factor Analysis: The Radex. In *Mathematical Thinking in the Social Sciences*; Lazarsfeld, P.F., Ed.; Free Press: New York, NY, USA, 1954; pp. 258–348.
12. Lee, K.C.; Chung, N. Empirical analysis of consumer reaction to the virtual reality shopping mall. *Comput. Hum. Behav.* **2008**, *24*, 88–104. [\[CrossRef\]](#)

13. Bruno, F.; Mattanò, R.M.; Muzzupappa, M.; Pina, M. Design for Usability in virtual environment. In Proceedings of the ICED 2007, the 16th International Conference on Engineering and Design, Paris, France, 28–31 August 2007; pp. 1–12.
14. Stadler, S.; Cornet, H.; Huang, D.; Frenkler, F. Designing Tomorrow’s Human-Machine Interfaces in Autonomous Vehicles: An Exploratory Study in Virtual Reality. In *Augmented and Virtual Reality*; Tom Dieck, M.C., Jung, T.H., Eds.; Springer Nature: Cham, Switzerland, 2019.
15. Carlsson, M.; Sonesson, T. *Using Virtual Reality in an Automotive User Experience Development Process*; Chalmers University of Technology: Gothenburg, Sweden, 2017.
16. Stadler, S.; Cornet, H.; Novaes Theoto, T.; Frenkler, F. A Tool, not a Toy: Using Virtual Reality to Evaluate the Communication between Autonomous Vehicles and Pedestrians. In *Augmented Reality and Virtual Reality*; Tom Dieck, M.C., Jung, T.H., Eds.; Springer Nature: Cham, Switzerland, 2019.
17. Mäkelä, V.; Radiah, R.; Alsherif, S.; Khamis, M.; Xiao, C.; Borchert, L.; Schmidt, A.; Alt, F. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–15. [[CrossRef](#)]
18. Gabbard, J.L.; Hix, D.; Edward Swan II, J. User-Centered Design and Evaluation of Virtual Environments. In *IEEE Computer Graphics and Applications*; IEEE: New York, NY, USA, 1999; pp. 51–59.
19. Murtza, R.; Monroe, S.; Youmans, R.J. Heuristic evaluation for virtual reality systems. In *Proceedings of the Human Factors and Ergonomics Society*; Sage Publishing: London, UK, 2017; pp. 2067–2071. [[CrossRef](#)]
20. Wang, W.; Guo, J.L.C.; Cheng, J. Usability of virtual reality application through the lens of the user community: A case study. In *CHI EA’19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2019. [[CrossRef](#)]
21. Silvennoinen, M.; Kuparinen, L. Usability challenges in surgical simulator training. In Proceedings of the International Conference on Information Technology Interfaces (ITI 2019), Cavtat, Croatia, 22–25 June 2009; pp. 455–460. [[CrossRef](#)]
22. Zhang, X.; Simeone, A.L. Using heuristic evaluation in immersive virtual reality evaluation. In *ACM International Conference Proceeding Series*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 223–225. [[CrossRef](#)]
23. Rusu, C.; Muñoz, R.; Roncagliolo, S.; Rudloff, S.; Rusu, V.; Figueroa, A. Usability Heuristics for Virtual Worlds. In *The Third International Conference on Advances in Future Internet*; IARIA XPS Press: Wilmington, DE, USA, 2011; pp. 16–19.
24. Billow, T.V.; Cottam, J.A. Exploring the Use of Heuristics for Evaluation of an Immersive Analytic System. 2017, pp. 1–5. Available online: [https://groups.inf.ed.ac.uk/vishub/immersiveanalytics/papers/IA\\_1375-paper.pdf](https://groups.inf.ed.ac.uk/vishub/immersiveanalytics/papers/IA_1375-paper.pdf) (accessed on 19 December 2022).
25. Stadler, S. The Integration of Virtual Reality into the Design Process. Ph.D. Thesis, Technical University of Munich, Munich, Germany, 2021.
26. Labrie, A.; Cheng, J. Adapting Usability Heuristics to the Context of Mobile Augmented Reality. In *UIST 2020—Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 4–6. [[CrossRef](#)]
27. Sari, R.N.; Sri Hayati, R.; Fujiati; Rahayu, S.L. Heuristic Evaluation in Mobile Augmented Reality Applications in Designing Houses. In Proceedings of the 2020 8th International Conference on Cyber and IT Service Management, CITSM 2020, Pangkal, Indonesia, 23–24 October 2020; IEEE: New York, NY, USA, 2020; pp. 1–5. [[CrossRef](#)]
28. Nielsen, J. 10 Usability Heuristics for User Interface Design. Available online: <https://www.nngroup.com/articles/ten-usability-heuristics/> (accessed on 31 January 2023).
29. Lewis, C. Using the ‘think-aloud’ method in cognitive interface design. In *IBM Research Report RC 9265, 2/17/82*; IBM Thomas J. Watson Research Center: New York, NY, USA, 1982.
30. Deb, S.; Carruth, D.W.; Sween, R.; Strawderman, L.; Garrison, T.M. Efficacy of virtual reality in pedestrian safety research. *Appl. Ergon.* **2017**, *65*, 449–460. [[CrossRef](#)]
31. Stadler, S.; Cornet, H.; Frenkler, F. Towards user acceptance of autonomous vehicles: A virtual reality study on human-machine interfaces. *Int. J. Technol. Mark.* **2019**, *13*, 325–353. [[CrossRef](#)]
32. Pillai, A. Virtual Reality based Study to Analyse Pedestrian Attitude towards Autonomous Vehicles. Master’s Thesis, Aalto University, Espoo, Finland, 2017.
33. Kuliga, S.F.; Thrash, T.; Dalton, R.C.; Hölscher, C. Virtual reality as an empirical research tool—Exploring user experience in a real building and a corresponding virtual model. *Comput. Environ. Urban Syst.* **2015**, *54*, 363–375. [[CrossRef](#)]
34. LaViola, J.J. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bull.* **2000**, *32*, 47–56. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.