Systematic Review

Similarities and Differences between Immersive Virtual Reality, Real World, and Computer Screens: A Systematic Scoping Review in Human Behavior Studies

Daniel Hepperle [1,2,*] and Matthias Wölfel [1,2]

1   Institute for Intelligent Interaction and Immersive Experience, Karlsruhe University of Applied Sciences, 76133 Karlsruhe, Germany; matthias.woelfel@h-ka.de
2   Media Use Research Group, Faculty of Business, Economics and Social Sciences, University of Hohenheim, 70599 Stuttgart, Germany
*   Correspondence: daniel.hepperle@h-ka.de

Abstract: In the broader field of human behavior studies, there are several trade-offs for on-site experiments. To be tied to a specific location can limit both the availability and diversity of participants. However, current and future technological advances make it possible to replicate real-world scenarios in a virtual environment up to a certain level of detail. How these differences add up and affect the cross-media validity of findings remains a topic of debate. How a virtual world is accessed, through a computer screen or a head-mounted display, may have a significant impact. Not surprisingly, the literature has presented various comparisons. However, while previous research has compared the different devices for a specific research question, a systematic review is lacking. To fill this gap, we conducted this review. We identified 1083 articles in accordance with the PRISMA guidelines. Following screening, 56 articles remained and were compared for a qualitative synthesis to provide the reader a summary of current research on the differences between head-mounted displays (HMDs), computer screens, and the real world. Overall, the data show that virtual worlds presented in an HMD are more similar to real-world situations than to computer screens. This supports the thesis that HMDs are more suitable than computer screens for conducting experiments in the field of human behavioral studies.

Keywords: scoping review; human behavioral studies; comparison; immersive virtual reality; head-mounted displays; computer screen; real world

1. Introduction

To date, numerous experiments in the broader field of human behavior research have been and are being conducted in virtual environments. Virtual environments are used to overcome several trade-offs observed in on-site experiments. For example, to be tied to a specific location can limit both the availability and diversity of participants. Although tools, such as Amazon's Mechanical Turk (https://www.mturk.com/, accessed on 14 April 2023), PsyToolkit (https://www.psytoolkit.org/, accessed on 14 April 2023), or Eprime 3.0 (https://pstnet.com/, accessed on 14 April 2023), promise to make it easy to design experiments and collect data, they are mostly limited to content such as text, video, or images. Realistic 3D environments can overcome these limitations and can be used either on a *computer screen* or within a *head-mounted display* (HMD). Whereas the former provides a monoscopic display, the latter provides a stereoscopic display with a large field of view and content that adapts to the position of the head. In both cases, the term virtual reality (VR) is often used. To distinguish between the output devices (which also determines the type of input), "immersive" is often added in the case of HMDs. The higher degrees of freedom are combined with a stereoscopic viewport, as offered by HMD VR, and can make someone feel as if they are present somewhere else [1]. Combined with a well-implemented 3D world, it

can be so convincing that users forget the physical space they are in. Even though the use of HMD VR seems promising, it is important to note that HMD VR, like displays, has several limitations. These may be due to technical limitations (e.g., display resolution, refresh rate, field of view), or conflicting parameters (e.g., vergence–accommodation conflict [2], visuo-proprioceptive conflict [3]) that can lead to different unwanted effects in the way VR is perceived (e.g., uncanny valley effect [4], color perception [5], or suspension of disbelief). Although some of these effects and limitations are exclusive to HMDs, others are shared with more established output devices such as screens or *cave automatic virtual environments* (CAVEs). The CAVE is a cube with display screen faces surrounding a viewer [6]. In some cases, CAVEs may be a valid alternative to other entities; however, it is not included in this review due to the low number of results.

Although there may be a high *internal validity* (results that hold true within the environment), *cross-media validity* (results that hold true in other environments) cannot be assumed for experiments conducted in a virtual environment. This is especially critical when applying findings from a virtual environment to the real world. In this paper, we provide a systematic review of the evidence by comparing the results between HMD VR and the real world, and between HMD VR and screens.

Although it is obvious why comparing HMD VR to the real world is important for human behavioral studies, the rationale for also comparing it to screens may not be immediately obvious. Many of the advantages of using virtual environments in human behavioral research apply to both types of output devices. However, designing, setting up, and conducting HMD VR studies is significantly more challenging than conducting VR studies with a screen. Therefore, evidence on how HMD VR differs needs to be collected to provide a basis for deciding which setup (HMD, screen, or real world) to use. In addition, the collected information can serve as a valuable reference for researchers and developers who want to optimize their virtual environments for different applications and use cases.

In 2003, Frank Biocca framed the notion of presence in immersive VR as *"how the mind 'perceives' reality, not reality itself; not physics but psychology; the extended mind, the place where experience, technology, and psychology meet"* [7]. This suggests that immersive VR has less to do with a physical setup and more to do with the mind. However, in order to realize the full potential of HMD VR, it is necessary to understand the technical aspects in order to know how to plan and build these experiences to achieve the desired effect.

All individual findings are listed and categorized into major categories (e.g., interaction or perception) and subcategories (e.g., efficiency or presence) for ease of reference. In addition, all findings are compared to each other to provide an overview of similar and contrary results. For each paper, we also collected information about the study population such as the number of participants, gender, and age. Regarding the research methodology, we collected information about the questionnaires used, the study design (within groups; between groups), and the software and hardware used to conduct the study. In addition, the goal was to identify any gaps so that new research activities can be positioned accordingly. This is essential to understand this new technology in such a way that the specific needs and possibilities of HMD VR can be addressed and compared to other research environments.

We are interested in understanding if and how HMD VR can be used to conduct research on human behavior. However, this is only possible if the conclusions drawn in (immersive) VR can be applied to the real world. By posing and answering the following research questions, we aim to highlight the similarities and differences that require further attention when conducting research in virtual environments with the goal of applying these findings to the real world:

RQ1: *"What are the main differences between HMD VR, screen-based VR, and the real world mentioned in the current literature?"*

RQ2: *"What are the expected consequences of these differences?"*

RQ3: *"How extensive are these differences?"*

Our initial goal was to provide evidence specifically for human behavioral studies, yet the findings are not limited to this specific research discipline. Knowledge of similarities and differences between environments is helpful in all cases where knowledge, insights, etc., need to be transferred from one to another. Examples include education and training (e.g., surgical training [8], pilot training [9], safety training [10]), sports [11] and physiotherapy [12], human factors engineering [13], or for exposure therapies to treat anxiety or similar issues [14,15].

2. Related Work and Theoretical Foundation

In the late 90s and early 2000s, papers were published attempting to examine the differences between HMD VR and other technologies as well as actual reality. For example, Yoon et al. [16] published a paper in which they examined spatial perception using an HMD VR and compared it to how it is perceived in the real world. They found that spatial perception in general did not differ from real-world perception, but height estimates did. Although research has examined how HMD VR compares to other entities within a single paper or study, to the best of our knowledge, not much work has been carried out to provide a comprehensive overview. Santos et al. [17] took the first step in mapping the research landscape in this context more than ten years ago, most likely as a byproduct of their original intent to measure navigation performance between desktop systems and HMD VR. As the technology has advanced tremendously in the last 15 to 20 years, one might wonder if the outdated findings are still valid. For example, the HMD used by Yoon et al., the *V8* (http://www.virtualresearch.com/products/v8.htm, accessed on 14 April 2023) from Virtual Research Systems, Inc. came with a 60° diagonal *field of view* (FoV), while current consumer-grade headsets offer a diagonal FoV of 110° (HTC Vive Pro) and in some cases as much as 170° (Pimax 5K+). Fundamental research work, such as Milgram and Kishino's Virtual Reality Continuum introduced in 1994 [18], provides definitions along the reality spectrum. Furthermore, systematic reviews examining the differences between virtual and augmented reality exist [19]. Despite these existing studies, our research seeks to address a notable gap: there has been a lack of exploration regarding the distinctions among real-world entities, immersive virtual reality, and screen environments. This became evident during our exhaustive search, where we were unable to identify any systematic scoping reviews focusing on this particular topic. In order to conduct a systematic scoping review that can contribute to scientific progress by providing a well-researched and aggregated overview of the current research landscape, we followed the guidelines suggested by [20]. They defined the goal of a scoping review as "*to determine what range of evidence (quantitative and/or qualitative) is available on a topic and to represent this evidence visually as a mapping or charting of the located data.*" As some procedures in a systematic scoping review are similar or adopted from a systematic review, we also used the PRISMA guidelines given in [21] for the orientation and structure of this work.

2.1. Categories

To categorize the fundamental aspects of VR, we propose *perception*, *interaction*, and *sensing and reconstruction of reality* as the main categories.

2.1.1. Category I—Perception

Perception can be described as the entirety of impressions that are received by our senses. In general, immersive technologies are able to simulate these impressions to a certain degree. The more and the better this is simulated or faked, the less a person is able to distinguish between the technology and the actual real world.

2.1.2. Category II—Interaction

Interaction is important for any technical system with the goal that users can not only perceive information, but also manipulate it. This cooperation between a technical system and the user is necessary in cases where there is no predefined linear narrative, but where

the content can be modified by the user. This category includes results in the area of efficiency (i.e., time to task completion), usability, or workload. More specifically, issues such as object manipulation, navigation, or ease of use fall into this category.

2.1.3. Category III—Sensing and Reconstructing Reality

No matter how hard one tries, it is impossible to completely detach from reality. Whether it is something as basic as moving furniture in your HMD VR installation, or something more subtle, such as temperature or smell, users are always affected by their physical surroundings. In order to achieve an optimal mapping, it is essential to sense the real world and to reconstruct it within the virtual world in some way. This does not necessarily have to be a 1 to 1 replica, but can vary to some degree [22–24]. It does not have to be a static reconstruction of the environment. The use of cameras or other sensor systems combined with machine learning algorithms allows you to sense or scan your environment in real time so that you can implement facial expressions or project full avatar movements and corresponding textures into the HMD VR experience [25,26].

2.1.4. Subcategories

For each individual article collected in the screening process, a category was noted by the authors without restriction. These were grouped into more similar categories. Subcategories are not limited to a specific main category but can also fall into two main categories. For example, in efficiency as a subcategory can be perceived efficiency (i.e., result no. 39: "sig. higher felt individual performance in VR") and therefore would fit as a subcategory of perception, but it can also be a subcategory of interaction (i.e., result no. 16 "sig. faster in time to task completion").

2.2. Compared Settings

In this work, we focus on the following *intervention types* (IVT) on which the comparison is based: VR (screen), immersive VR (HMD), and real world. Contrary to what was stated in the preregistration, CAVE is no longer considered due to the low number of results. For all IVT, it is important to note that we do not differentiate between the specific content, e.g., if it is presented as an image or as a 3D point cloud, we only focus on the device type. We refer to

- *Screen:* as monoscopic displays in all different sizes as they are commonly used on a PC or tablet;
- *HMD VR:* as all kinds of head mounted displays that visually isolate the user from the environment. Content can range from interactive stereoscopic 3D computer graphics to 360° video or photos; and
- *Real world:* as the world that seems to exist.

3. Methodology

We conducted a scoping review to map the research landscape of the unique characteristics of HMD VR compared to established technologies and procedures. A VR-related example of a scoping review would be the work of [27], in which they provided an overview of work dealing with VR technology in the assessment and treatment of psychosis. This work is based on the recommended steps as suggested by [20,28] and the PRISMA (preferred reporting items for systematic reviews and meta-analyses by [21]) guidelines for reporting.

3.1. Risk of Bias

As with all systematic scoping reviews, there is the problem of publication bias (also known as the file drawer effect) [29], which, in simple terms, states that studies are less likely to be published if no significant effect is found. This is of particular importance in this study, since we are interested in *differences* as well as whether there are *no differences* to be expected. No differences are of particular interest with regard to real-world comparisons,

since a finding of no differences would mean that HMD VR could be a valid substitute for a real-world study.

3.2. Query Development and Search

Creating the search query was an iterative process with several loops to define the final search string. The query can be seen in Figure 1. For each database (see Table 1), all BibTeX entries, including abstracts, were exported and imported into a locally installed version of the open source systematic literature review tool parsif.al, where all duplicates were removed [30]. Following the work of [31], in which they evaluated the respective qualities of 28 academic search systems, the ones listed in Table 1 were selected.

```
("Virtual Reality" OR "HMD" OR "VR" OR "Head
Mounted Display" OR "3D" OR "Stereo 3D")
AND (("Display" OR "Monitor" OR "2D" OR
"Screen") OR ("CAVE") OR ("Real World"))
AND ("Differences" OR "Similarities" OR "Com-
parison" OR "Correlation")
AND ("User Study" OR "Evaluation")
```

Figure 1. Search query developed and used by the authors.

Table 1. Overview of the search systems used.

Search System	Url
ACM Digital Library	https://dl.acm.org/
Arxiv only 2020	https://arxiv.org/
IEEE Xplore	https://ieeexplore.ieee.org/
Ovid	https://ovidsp.dc1.ovid.com/ovid-a/ovidweb.cgi [1]
Scopus	https://www.scopus.com/home.uri
Wiley Online Library	https://onlinelibrary.wiley.com/

[1] Login required

The criteria for selecting our search engines were as follows:

- The search engine must be thematically relevant. We included search systems from the fields of computer sciences, social psychological studies, behavioral studies, health sciences, and multidisciplinary studies with a focus on computer science and medicine;
- All search systems need to be able to make use of boolean operators in search strings (we used only OR; AND; NOT) [32]; and
- Are capable of more complex search terms (e.g., are able to make use of more than seven boolean separated search strings).

Even though a detailed pre-selection has been made, there are still some hurdles to overcome between the different search engines, especially syntactical ones. For example, we decided to search only within the abstracts of the available research articles, which in some cases had to be checked as additional criteria and sometimes could be implemented within the search string. The same happened when we tried to limit the search results to results that published after 2013. We considered articles published after 2013 because that was the year that Oculus Rift DK1 was shipped [33].

3.3. Preregistration

The study has been pre-registered and is available online at the Open Science Framework (URL to pre-registration at the Open Science Foundation: https://osf.io/gmfns/?view_only=274f99fd32384f42a877526134227337, accessed on 14 April 2023). The following

derivations of the pre-registration have been made. The entry fields used for collecting the data were supplemented by the entries: "VR hardware used, other hardware, comments, software used, and "what is being compared" to collect more data that might be of interest. The selection criterion "large screen" was not found. Due to the small number of results (1 each), the intervention types CAVE and Audio are not discussed in this paper.

4. Screening, Selection, and Assignment Procedure

The process for selecting and rejecting studies can be separated into the following four stages:

Stage 1: Immediately after searching the respective databases, the results were filtered by year (published after 2013) if this was not possible with the search string.

Stage 2: The abstracts of each record were screened according to the following selection and deselection criteria: All articles related to the IVT we defined were selected (see Section 2.2 for definitions). If any of the research articles used augmented reality (AR) instead of VR, it was rejected. If an article compared both VR and AR, it was not rejected. It is also a balancing act to make the search query as broad as possible and as narrow as necessary. As a result, many articles were found that made a comparison *with* an HMD VR environment and not *with* another IVT. These articles were also rejected. If something other than the above IVT was compared, it would also be rejected. Languages other than English were rejected. Articles that did not adequately document their research or explain the reasoning behind their conclusions were rejected based on the "unsound methods" rejection criterion.

Stage 3: The accessibility of all papers was checked. At this stage, we had to reject two more papers because they were not accessible.

Stage 4: All remaining papers were screened according to the data extraction suggested by [20] and selected or deselected accordingly. The following information was entered into the data extraction form for each paper selected after screening the abstract. Here a derivation to the pre-registration has been made. The entry fields after field 13."*Bibtex entry*" were added because they were mentioned in many study descriptions and are, in our opinion, a valuable addition to the mapping of the research landscape:

1. Author(s)
2. Year of publication
3. Source of origin / country (if accessible)
4. Aims/purpose
5. Study population
6. Sample Size
7. Methodology
8. Intervention Type (IVT) / Tech. Used
9. Concept
10. Duration of intervention
11. How outcomes are measured
12. Key findings
13. Bibtex entry
14. What is compared
15. VR hardware used
16. Other hardware
17. Annotations
18. Software used

Most of the listed items are self-explanatory, but for a better understanding of how we classified the results, it is necessary to define the type of data collected under the item "What is compared". Here the authors noted the specific topics that were compared in

the article. Later, this information was used as described in Section 4.1 to derive a main category and subcategory that best fit the topic.

4.1. Postprocessing

All findings were assigned to one of the three (main) categories *perception*, *interaction*, and *sensing and reconstructing reality*. The assignment of each subcategory to the appropriate category was carried out by two people independently. The cases that differed between the two judgments were discussed again until a decision could be made. When necessary, a subcategory, such as efficiency, was assigned to more than one main category.

4.2. Prisma Flow Diagram

The flowchart shows the exact number of records found, selected, rejected, or removed due to duplication during the process (see Figure 2). The distribution of the results from the different search engines are as follows: Scopus 64%, Wiley Online Library 14%, IEE 13%, OVID 4%, ACM 4% (numbers are rounded for better readability). As can be seen, two additional records were selected from other articles because the cited references seemed to be a valuable contribution to this review. Following identification, screening, and eligibility checking, 56 articles were included in this scoping review. Of the 56 articles finally included in this scoping review, nearly 56% compared immersive VR (HMD) to VR (screen) and 44% compared immersive VR to the real world. Only three articles compared all three settings, immersive VR, VR, and the real world.
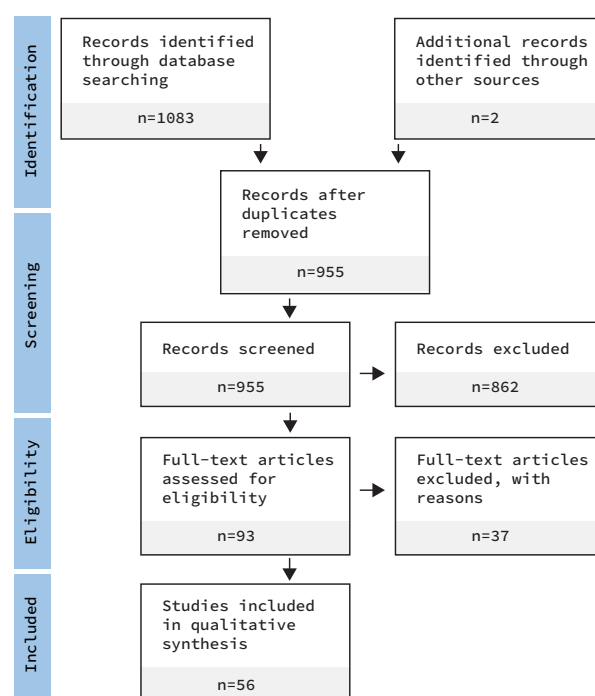
Figure 2. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram for the scoping review process [21].

5. Results

Most of the 56 studies included in the synthesis report more than one finding, resulting in a total of 163 findings, some of which can be compared to other findings. Sometimes, a single research paper answered more than one question. So, we ended up with more findings than the total number of papers we examined. Each finding has been assigned one category and one subcategory.

To improve comprehension, individual findings in the tables are categorized and visually distinguished using icons that vary in shape and color. This helps to indicate whether the result is from the perspective of the HMD VR:

▲   *Advantageous* in relation to the screen or real world;

▼   *Disadvantageous* in relation to the screen or real world;

▶   *Similar* in relation to the screen or real world if there is no significant difference; and

■   *Undecided* if no clear tendency can be inferred, but there is a significant difference.

Table 2 shows the number of results for each main- and subcategory. For a better understanding, the listings in the table are to be read from the perspective of HMD VR. For example, in the first row, in the *interaction* category, the results related to the *efficiency* subcategory, in comparison to the real world, show that there are three results in favor of HMD VR. Ten results show no difference between HMD VR and the real world, zero results are undecided, and four results are against HMD VR. Similarly, for screen, 10 findings are favorable to HMD VR, seven show no significant difference between HMD VR and screen, zero are undecided, and in two findings, it is stated that the HMD VR-related task was less efficient than in the screen environment. For each subcategory, the results are summed and the distribution between favorable, similar, undecided, and unfavorable findings is expressed as a percentage.

5.1. Hard- and Software Setup

To provide a comprehensive overview of the hardware and software used, the available data from the papers are summarized as follows: an Oculus device was used 42 times (we do not count the Samsung Gear VR as Oculus, even though it is co-developed by Oculus), a device from the HTC Vive family was used 33 times, the Samsung Gear VR was used four times, the Google Daydream was used one time, and the Pimax 5k HMD was used one time. In 17 cases, the hardware was not specified. In the case of software, we can see that, if specified, 92% (44) of the studies were created using the Unity game engine, while only 8% (4) used Unreal (please note that the numbers may not add up as expected, since in some works, two different HMDs were used).

5.2. Study Population and Duration

As shown in Table 3, the majority of participants where gender was specified (n: 1051; f: 405; m: 644; d: 2) were male (61.3%). The number of participants ranged from as few as three to as many as 200 (gender not specified). The age of the participants ranged from 17 to 85 years. In terms of study duration, one outlier included an observation period of a full day (8 h). Excluding this outlier, the average study duration was 41 minutes (SD: 27).

Table 2. The table shows the number of results distributed among each subcategory comparing HMD VR to the real world and screen.

Category		HMD VR in Comparison to:							
		Real World				Screen			
	Sub-Cat.	▲	▶	■	▼	▲	▶	■	▼
Interaction	Efficiency	3	10	0	4	10	7	0	2
	Interaction	0	3	0	0	2	0	0	0
	Overview	0	0	0	0	2	0	0	0
	Physical Demand	0	0	0	0	0	0	0	1
	Simulator Sick.	0	0	0	1	0	0	0	0
	Usability	0	1	0	1	0	3	0	2
	Usefulness	0	1	0	0	0	0	0	0
	User Experience	0	1	0	0	0	0	0	0
	Workload	1	2	0	3	0	1	0	0
	Σ	4	18	0	9	14	11	0	5
	%	13	58	0	29	47	38	0	16

Table 2. *Cont.*

Category	Sub-Cat.	Real World ▲	Real World ▶	Real World ■	Real World ▼	Screen ▲	Screen ▶	Screen ■	Screen ▼
Perception	Aesthetics	0	1	0	0	0	0	0	0
	Accuracy	0	0	0	0	1	0	0	0
	Color	0	0	0	0	0	0	2	0
	Efficiency	0	0	0	0	3	0	1	0
	Emotions	0	1	1	0	0	0	0	0
	Engagement	0	4	0	1	3	3	0	0
	Experience	0	0	0	0	1	0	0	0
	Frustration	0	0	0	0	3	0	0	1
	Immersion	0	0	0	0	3	0	0	1
	Learning	0	2	0	0	2	2	0	3
	Motion Sickness	0	1	0	1	0	0	0	0
	Perception	0	0	0	0	0	1	0	0
	Presence	1	1	0	0	6	2	0	0
	Qual. of Exp.	0	0	0	0	0	0	0	1
	Realism	0	1	0	1	1	0	0	0
	Satisfaction	0	0	0	0	2	0	0	0
	Simulator Sickness	0	0	0	0	0	0	0	1
	Spatial Perception	0	0	0	0	1	0	0	0
	Workload	1	0	0	0	4	2	0	4
	Σ	2	11	1	3	33	10	3	11
	%	12	65	6	18	58	18	5	19
Sensing and Reconstr.	Accuracy	0	0	0	0	1	1	1	0
	Autonomy	0	0	0	0	1	0	0	0
	Efficiency	0	0	0	0	0	1	0	0
	Flexibility	1	0	0	0	0	0	0	0
	Haptics	0	1	0	0	0	0	0	0
	Interaction	1	0	0	0	0	0	0	0
	Learning	1	0	1	0	0	0	0	0
	Locomotion	0	0	0	1	0	0	1	0
	Overview	0	0	0	0	1	0	0	0
	Physi. Response	0	0	0	0	0	1	0	0
	Realism	0	1	0	0	0	0	0	0
	Reconstruction	1	0	1	0	0	0	0	0
	Spatial Perception	0	2	0	3	0	1	0	1
	Transferability	0	0	1	0	0	0	0	0
	Usability	0	0	0	1	0	0	0	0
	Σ	4	4	3	5	3	4	2	1
	%	25	25	19	31	30	40	20	10

[▲]: No. of results that are advantageous towards HMD VR; [▶]: No. of similar results (no sig. difference found);
[■]: No. of indecisive results—not able to infer a tendency; [▼]: No. of results in which HMD VR is a drawback.

Table 3. Overview of the study population from the accepted articles. Some studies did not mention age and information about participants; therefore, these could not be taken into account in the calculation.

	Male	Female	Diverse	Not Defined	Age Min–Max
Average	16.95	10.66	0.05	27.61	22–39
SD	13.45	11.35	0.21	4.14	8–18
n	38	38	2	22	25
Σ	644	405	2	574	x

5.3. Questionnaires Used

The questionnaires used are rather fragmented. Nevertheless, almost 70% (n: 39) of the 56 papers used a questionnaire. All questionnaires that were not developed by the authors themselves and cited accordingly can be found in Table 4. The questionnaires used in the 56 works assess aspects such as task load, presence, usability, user experience, engagement, and simulator sickness. The most commonly used questionnaire is the NASA task load index (TLX) [34], indicating that workload or cognitive load is an important factor that was examined in these studies. Other commonly used questionnaires, such as Witmer and Singer's presence questionnaire [35] and the system usability scale (SUS) [36], indicate that researchers are also interested in understanding the sense of presence and usability of the systems under study.

Table 4. Number of questionnaire usages found within the 56 articles that are included in this review. No.: number of usages counted over all articles collected. Origin: inventor of the questionnaire. Used In: articles in which the questionnaires are used.

No.	Questionnaire	Origin	Used In
7	Task Load Index (TLX) by NASA	[34]	[37–43]
4	System Usability Scale (SUS)	[36]	[38,44–46]
2	Witmer and Singer's Presence Questionnaire	[35]	[42,47]
2	User Experience Questionnaire (UEQ)	[48]	[49,50]
2	IBM CSUQ System Usability	[51]	[42,47]
1	IGroup Presence Questionnaire (IPQ)	[52]	[39,53]
1	After-Scenario Questionnaire (Satisfaction)	[54]	[53]
1	Immersive Experience Questionnaire (IEQ)	[55]	[56]
1	ITC-Sense of Presence Inventory	[57]	[50]
1	Player Experience of Need Satisfaction (PENS) Questionnaire	[58]	[59]
1	Self-Assessment Manikin (SAM)	[60]	[61]
1	Simulator Sickness Questionnaire	[62]	[63]
1	Temple Presence Inventory (TPI)	[64]	[65]
1	Intrinsic Motivation Inventory (IMI)	[66]	[38]
1	Virtual Reality Sickness Questionnaire (VRSQ)	[67]	[38]
1	User Engagement Scale	[68]	[69]
1	Satisfaction and Self-Confidence in Learning (SSCL)	[70]	[71]

5.4. Study Design

In the case of the study design, a between-group design was used 36 times and a within-group design was used 40 times. The almost even distribution shows that research on HMD VR is interested in individual differences or changes within individuals over time, as well as comparisons with different groups. In addition, within-group designs can be useful in situations where it is difficult or impossible to recruit enough participants from a particular group.

5.5. Mapping the Field

In summary, we obtained an overview on the most studied topic. We can see that more than 61.3% (n: 100) of the results refer to a comparison between HMD VR and the screen environment, while only 38.7% (n: 63) of the results compare HMD VR with a real world scenario. As seen in Figure 3 most of the research carried out on the HMD VR × screen setting falls into the *perception* category (n: 57) next to *interaction* (n: 32), and 10 of the findings relate to the *sensing and reconstructing reality* category. However, in the HMD VR × real comparison, the results are more evenly distributed across the categories. Similar to HMD VR × screen, *sensing and reconstructing reality* is the category with the fewest results (n: 16). However, the number of results for the *interaction* and *perception* categories are reversed. This means that for *interaction*, they have the most (n: 31), and for *perception*, the second most (n: 17) results for HMD vs. real world.

In both comparison settings, the most researched subcategory is *efficiency* with 19 results for HMD VR × screen and 17 results for HMD VR × real world. Other highly investigated subcategories for HMD VR × screen are *workload* with 11 results (in the main category interaction, 10 in perception), *presence* with eight, and *learning* with seven. For HMD VR × real world they are *workload* with seven results (six for interaction and one for perception), *engagement* has five results, and *spatial perception* has five results. In addition, the range of questionnaires used in these papers highlights the complex nature of studying immersive virtual reality, computer screens, and the real world, as well as the need for multiple instruments to capture the different dimensions of the user experience in these environments.
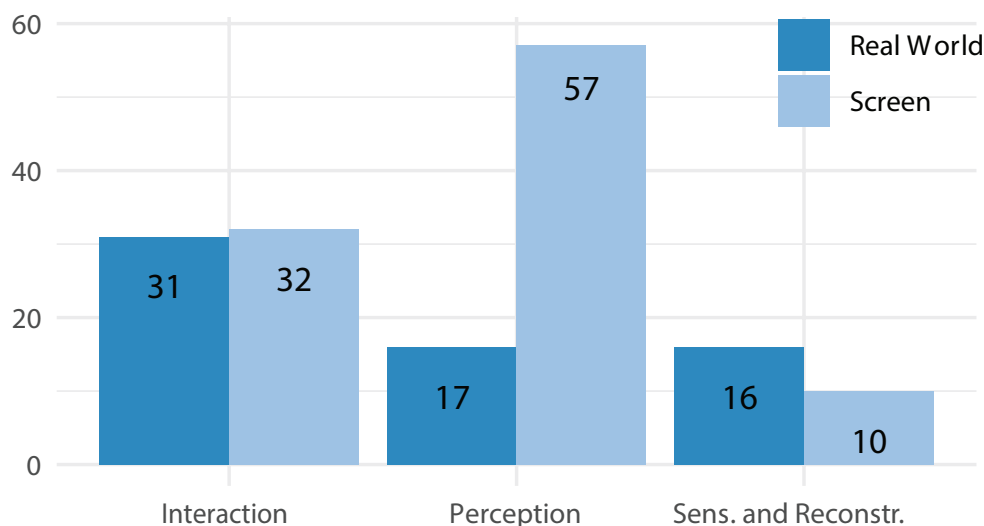

Figure 3. Number of results for the three main categories by intervention type.

5.6. Advantages and Disadvantages in General

One can conclude from the results that HMD VR is advantageous in 37% (n: 61; 51 × screen; 10 × real world) of the 163 results and disadvantageous in 21% (n: 34) of the cases (17 × screen; 16 × real world); 5.5% (n: 9) cases cannot be classified due to the nature of the findings (5 × screen; 4 × real world). For example, finding number 34 "sig. stronger fear" in HMD VR × real world may be positive if VR elicits higher emotional arousal, but fear may also be a disadvantage. The other 36% (n: 59) of findings (26 × screen; 33 × real world) are categorized as *no difference*. When assigned to the "no difference" category, no significant difference could be found between HMD VR and the compared entity. As a main finding, we observe a high number of *similarities* in the results for HMD VR × real world. More than 50% (n: 33) of the 63 results for VR × real world show no significant difference. This is of great interest because some differences are likely due to technical limitations that may be resolved in the future.

We argue that it is of the utmost importance to understand and evaluate the specific characteristics when looking first at the IVT, such as real world or screen, and then at the category. In some cases, a portion of the subcategories may be the same for another category. This is due to the nature of the research and sometimes comes to its limits in breaking results down. Here we invite the reader to take a closer look at the work in question, as it cannot be described in more detail within the scope of this work.

5.7. Advantages and Disadvantages per Category

Examining the advantages, disadvantages, and similarities per category allows us to gain more information about **RQ1:** *"What are the main differences between HMD VR, screen-based VR, and the real world mentioned in the current literature?"*. Note that in this section. we are interested in both differences (advantages and disadvantages) and similarities. If the mentioned percentages do not add up to 100%, this is either due to rounding or to the

fact that for this category, there are results categorized as "indecisive", which cannot be assigned to either differences or similarities.

5.7.1. Interaction Category

When comparing HMD VR × real world, 13% (n: 4) of the results showed HMD VR to be advantageous, and 29% (n: 9) of the results showed HMD VR to be disadvantageous. Almost 60% (n: 18) of the results showed no significant difference. This means that the percentage of differences is 42% (n: 13) compared to the percentage of similarities 58% (n: 18). Overall, we found more similarities than differences for HMD VR × real world in the *interaction* category. Moreover, when comparing HMD VR × screen, we found 62% of differences (n: 20) and 37.5% of similarities (n: 12). Therefore, the results suggest that HMD VR is more similar to real-world environments than to screen environments in terms of the *interaction* category.

5.7.2. Perception Category

In the *perception* category, we find similar tendencies as for *interaction*. We observe that 64% (n: 11) of the results are similar between HMD VR × real world and only 29% (n: 5) are different between the two entities. For HMD VR × screen, the percentage of differences is 77% (n: 44) and the percentage of similarities is 18% (n: 10). Therefore, in the *perception* category, the results suggest that HMD VR is more similar to the real world than to screen environments.

5.7.3. Sensing and Reconstructing Reality Category

Sensing and reconstructing reality is the only category that shows more differences at 56% (n: 9) than similarities at 25% (n: 4) for HMD VR × real world. For the HMD VR × screen comparison, we see equally distributed differences at 44% (n: 4) and similarities at 44% (n: 4). Furthermore, it is worth mentioning that we found only one case in which HMD VR was worse than screen for the *sensing and reconstructing reality category*.

5.8. Possible Consequences

Answering **RQ2:** *"What are the consequences of these differences?"*, we observe when comparing HMD VR with findings from the real world and from the screen that there are *more similarities between HMD VR × real world* for the interaction and perception subcategories than we find similarities between HMD VR × screen. For sensing and reconstructing reality, the *similarities and differences are more evenly distributed*. Overall, the results indicate that HMD VR environments tend to be more similar to real-world environments than screen-based environments in terms of interaction and perception. This may be useful information for designers and researchers looking to create more immersive and realistic virtual experiences.

When considering **RQ3:** *"How elaborate are these differences?"*, it seems that the similarities outweigh the differences between HMD VR × real world. This suggests the potential for using HMD VR as a platform for experimentation, as noted in Section 1. For HMD VR × screen, the differences outweigh the similarities, which could be an important sign in cases where existing studies on a screen could be transferred to an HMD VR scenario. However, it remains uncertain how detailed these differences are. A particular drawback in this regard is the measured effect. As mentioned above, some reported sample sizes are rather small, which is in line with the findings of [72]. In order to better understand how studies are conducted in this area, future work should consider effect sizes and study design.

5.9. Corresponding and Contradictory Findings

Table 5 lists all results comparing HMD VR with the real world results. Table 6 lists all results comparing HMD VR with the results obtained in a screen environment. The "Corr." and "Contr." columns group the results together and compare whether they are correlated or contradict each other.

For each finding, we list in columns 5 and 6 of Tables 5 and 6 the results that are related to each other either because they support the same hypothesis or because they present conflicting results. Since new results are usually easier to publish than successful and unsuccessful replications of an experiment, we have not found a 1 to 1 replication of an experiment that would prove a result to be more robust. To provide an overview, we believe it is useful to relate results that are in the same category and subcategory, as long as the detailed findings are thematically similar. We consider the analysis of corr. and condr. results to be an initial guide for future studies and to provide a brief overview of the current research situation, but a close examination of the respective work is required.

5.9.1. Single Findings HMD VR × Real World

To obtain a better understanding of corr. and contr. findings, we will take a closer look at each finding and list them accordingly. As mentioned before, it is of the utmost importance to take a closer look at the cited literature, as these are not generalizable results, but only specific cases in which this finding applies. For example, one of the studies examined forklift operator behavior and showed high correlations with the behavior observed in real-world situations.

Efficiency Most studies report no significant differences in task completion time (No. 3, 6, 9, 11), error rates (No. 12), or entry accuracy (13). Eye-gaze input (No. 15), felt individual performance (No. 4), and task-related focus (No. 5) are reported to be advantageous in HMD VR. HMD VR is disadvantageous as some studies found reaches to be less efficient (No. 7), higher time to task completion in VR (No. 8), slower object placement (No. 1), and slower touch input (No. 14).

Interaction Interaction skills show no significant difference (No. 18, 19) and similar qualitative feedback (No. 20) between VR and the real world.

Simulator Sickness Higher simulator sickness is reported in VR (No. 21).

Usability Usability results are mixed, with no significant differences found in some studies (No. 22) and lower scores for ease of use in VR in others (No. 23).

Usefulness VR-based aging simulation is found to have the same potential as real-world aging suits in terms of usefulness (No. 24).

User Experience No significant difference in user experience is reported between VR and the real world (No. 25).

Workload Workload results are mixed, with some studies reporting no significant differences in cognitive load (No. 26, 30) and others reporting higher mental demand (No. 27) and lower workload in VR (No. 31).

Aesthetics No difference in aesthetic preferences between VR and the real world (No. 32).

Emotions Emotion findings are mixed, with no significant difference between VR and video for most emotion arousal (No. 33) but stronger fear in VR (No. 34).

Engagement Engagement findings are varied, with no difference in engagement (No. 35), rapport (No. 37), co-presence (No. 38), and interpersonal trust (No. 39). Yet, one study reported lower engagement in VR (No. 36).

Learning No significant learning differences between learning (No. 40, 41, 52) but contradicting results exist (No. 51).

Motion Sickness More symptoms of "focus difficulty", "general discomfort", "nausea", and "headache" in VR (No. 42), but no difference in accommodation response (No. 43).

Presence Presence findings are mixed, with no significant difference in presence (No. 44) but a higher sense of presence in VR (No. 45).

Realism No significant differences between evaluation based on real user (supernumerary) in real world and avatars (No. 46), but lower natural feeling in VR (No. 47).

5.9.2. Single Findings HMD VR × Screen

The findings are similar to those in Section 5.9.1. A brief overview of the findings with more than one result will be discussed:

Efficiency With 10 results in favor for HMD VR, results in the efficiency subcategory shows a clear tendency towards HMD VR.

Overview Overview also leans towards VR, with results showing that data overview and data depiction (No. 22, 23) are more intuitive in VR.

Immersion, Experience Studies report higher immersion in VR (No. 50, 55, 56, 57) and lower frustration levels (No. 51), but also disadvantages such as a lower quality of experience (No. 75, 79) and a decrease in immersion at the narrative level (No. 58).

Learning Learning presents mixed results. Some studies suggest no significant differences in correct insights (No. 59), others suggest fewer correct insights in VR (No. 60). Others still report fewer deep insights from VR (No. 62), less learning in VR (No. 64), but also higher recall of information about tasks in VR (No. 63) and higher motivation in learning (No. 65).

Presence Presence in VR is generally found to be higher (No. 68, 69, 70, 71, 73, 74), although two studies report no significant difference (No. 67, 72) .

Satisfaction Data exploration is considered more satisfying in VR (No. 77) and VR is found to be more engaging (No. 78).

Workload Workload results are mixed, some studies report a lower workload in VR (No. 82, 84, 88), but others indicate higher cognitive load (No. 85, 86, 89).

Table 5. HMD VR compared to the real world.

	Sub-Category	Tend	Finding	No.	Corr.	Contra.	Reference
Interaction	Efficiency	▼	Sig. slower in object placement	1			[73]
	Efficiency	▶	VR based aging simulation has same potential as RR aging suits in terms of effectiveness	2			[74]
	Efficiency	▶	No sig. difference in time to task completion	3	9	8;29	[42]
	Efficiency	▲	Sig. higher felt individual performance in VR	4			[42]
	Efficiency	▲	Higher task-related focus in VR	5			[75]
	Efficiency	▶	No difference in task completion time when adding visuo-haptic feedback	6			[76]
	Efficiency	▼	Reaches were less efficient in the VR	7			[76]
	Efficiency	▼	Higher time to task completion in VR	8	29	9	[76]
	Efficiency	▶	No sig. difference in time to task completion	9	3	8;29	[77]
	Efficiency	▶	No sig. difference in score	10			[77]
	Efficiency	▶	No sig. difference in reading performance	11			[37]
	Efficiency	▶	No sig. difference in error rates	12			[37]
	Efficiency	▶	No sig. differences for entry accuracy	13			[78]
	Efficiency	▼	Sig. slower touch input in VR	14			[78]
	Efficiency	▲	Sig. faster eye-gaze input in VR	15			[78]
	Efficiency	▶	No sig. difference in finding an object	16			[73]
	Efficiency	▶	No sig. difference for grasping time and head movement	17			[73]
	Interaction	▶	No sig. difference in interaction skills	18	19;22	23	[79]

Table 5. *Cont.*

	Sub-Category	Tend	Finding	No.	Corr.	Contra.	Reference
Interaction	Interaction	►	Operation behavior of the same task in VE is highly correlated to that in RR (r > 0.90), which suggests VR successfully induces operation behavior, which is similar to the real operation behavior	19	18;22	23	[80]
	Interaction	►	Similar qualitative feedback in VR and real world condition	20			[78]
	Simulator Sickness	▼	Sig. higher simulator sickness	21			[37]
	Usability	►	No sig. diff in usability	22	18;19	23	[42]
	Usability	▼	Sig. lower score for ease of use	23		18;19;22	[49]
	Usefulness	►	VR-based aging simulation has same potential as RR aging suits in terms of usefulness	24			[74]
	User Exp.	►	No sig. difference in user experience	25			[50]
	Workload	►	No sig. difference in cognitive load	26			[77]
	Workload	▼	Sig. higher metal demand in VR	27			[37]
	Workload	▼	Sig. higher physical demand in VR	28			[37]
	Workload	▼	Sig. higher time to task completion	29	8	3;9	[37]
	Workload	►	No sig. difference in workload	30		31	[78]
	Workload	▲	Sig. lower workload in VR	31		30	[42]
Perception	Aesthetics	►	No difference in aesthetics preferences	32			[81]
	Emotions	►	No sig. difference between VR and video for each emotion arousal except fear	33	34		[82]
	Emotions	▇	Sig. stronger fear in VR	34	33		[82]
	Engagement	►	No difference in engagement	35		36	[65]
	Engagement	▼	Sig. lower engagement in VR	36	35		[69]
	Engagement	►	No difference in rapport	37			[69]
	Engagement	►	No difference in co-presence	38	44	45	[69]
	Engagement	►	No difference in interpersonal trust	39			[69]
	Learning	►	No learning differences between learning additive manufacturing in RR and VR	40	41;52	51	[83]
	Learning	►	No difference in learning success	41	40;52	51	[84]
	Motion Sick.	▼	Sig. more symptoms of "focus difficulty"; "general discomfort"; "nausea"; "headache" for VR	42			[85]
	Motion Sick.	►	No difference on accommodation response	43			[85]
	Presence	►	No sig. diff in presence	44	38	45	[42]
	Presence	▲	Higher sense of presence in VR	45		44	[65]
	Realism	►	No sig. differences between evaluation based on real user (supernumerary) in real world and avatars	46			[78]
	Realism	▼	Sig. lower natural feeling	47			[49]
Sens. and Recons.	Flexibility	▲	VR is advantageous compared to aging suits in terms of flexibility	48			[74]
	Haptics	►	No sig. difference in material identification when using the TAGlove compared to perceiving the real physical objects	49			[43]
	Interaction	▲	VR improves the external validity	50			[86]
	Learning	▲	VR kinesthetic experiences were more memorable and helped participants retain a larger number of words, despite any confounding elements that hindered their initial learning gain	51		40;41;52	[87]

Table 5. *Cont.*

	Sub-Category	Tend	Finding	No.	Corr.	Contra.	Reference
Sens. and Recons.	Learning	🟨	Participants first remembered sig. more words in the text-only conditon (RR); a week later, the amount of words remembered between text-only and VR with kinesthetic motion was equal	52	40;41		[87]
	Locomotion	▼	Significantly higher travel times in VR	53			[88]
	Realism	▶	No difference in realism	54			[65]
	Reconstruction	🟨	Transfer of motor skills from RR to VR not given	55			[89]
	Reconstruction	▲	VR studies completely support literature on real-life bike rides	56			[90]
	Spatial Perc.	▼	VR less accurate in distance estimation	57	58		[76]
	Spatial Perc.	▼	VR less correct in depth judgements	58	57		[76]
	Spatial Perc.	▶	No difference in distance estimation when adding visuo haptic feedback	59			[76]
	Spatial Perc.	▼	Sig. difference in behavior	60			[73]
	Spatial Perc.	▶	No sig. difference in distance traveled	61			[73]
	Transferability	🟨	Difference between therapist with experience in handling VR and therapists that had no prior experience; therapists with experience handled the patients the same as in conventional therapy whereas without experience they did not	62			[91]
	Usability	▼	VR generates fewer answers directly related with the mockup and more related to the surrounding	63			[86]

[▲]: HMD VR is advantageous [▼]: HMD VR is disadvantageous; [▶]: No difference between conditions; [🟨]: Indecisive; [**Corr.**] and [**Contr.**] describe if the finding confirms or is contradictory to other findings.

Table 6. HMD VR compared to the screen environment.

	Sub-Category	Tend	Finding	No.	Corr.	Contra.	Reference
Interaction	Efficiency	▼	Sig. slower filling out questionnaire in VR	1	5	3;7;10;12;17	[92]
	Efficiency	▲	Data exploration to be more successful in VR	2	11;22;23;77;4		[41]
	Efficiency	▶	No sig. difference in time to task completion	3	7;9;10;12	5;1;17	[42]
	Efficiency	▶	Data distinction similar	4			[40]
	Efficiency	▼	Time to task completion larger (slower) in VR	5	1	3;7;10;12;17	[93]
	Efficiency	▲	Performed better for design thinking tasks in VR	6			[94]
	Efficiency	▶	No difference in time to task completion	7	3;9;10;12	5;1;17	[95]
	Efficiency	▲	Reduced task error rate in VR	8			[95]
	Efficiency	▶	No differences in task completion time	9	3;7;10;12	5;1;17	[46]
	Efficiency	▶	No sig. difference in time to task completion	10	3;7;9;12	5;1;17	[47]
	Efficiency	▲	VR more efficient in data exploration	11	2;22;23;4		[96]
	Efficiency	▶	No sig. difference in time to task completion	12	3;7;9;10	5;1;17	[77]
	Efficiency	▶	No sig. difference in score	13			[77]
	Efficiency	▲	Sig. faster in annotation task	14			[97]
	Efficiency	▲	Sig. faster in counting	15			[97]
	Efficiency	▲	Sig. faster in time to task completion	16	17	3;5;7;9;10;12	[98]
	Efficiency	▲	Sig. faster in time to task completion	17	16	3;5;7;9;10;12	[38]
	Efficiency	▲	Sig. performance increase	18	14;15;19		[38]
	Efficiency	▲	Sig. faster in VR	19	14;15;18		[99]
	Interaction	▲	Interaction is more intuitive in VR	20	21;22;23	26	[40]
	Interaction	▲	Better interaction quality	21	20		[45]

Table 6. *Cont.*

	Sub-Category	Tend	Finding	No.	Corr.	Contra.	Reference
Interaction	Overview	▲	Data overview is easier in VR	22	20;23	26	[40]
	Overview	▲	Data depiction more intuitive in VR	23	20;22	26	[40]
	Phys. Demand	▼	VR data exploration required significantly more physical demand	24		82	[41]
	Usability	▶	No sig. difference in usability	25	26;27;32	28;29;31	[42]
	Usability	▶	No difference in intuitive controls	26	25;27;32	28;29;31	[59]
	Usability	▶	No sig. difference in usability	27	25;26;32	28;29;31	[47]
	Usability	▼	Sig. lower score in system usability scale questionnaire	28	29	25;26;27	[44]
	Usability	▼	VR is sig. harder to use	29	28	25;26;27;31	[100]
	Workload	▶	No sig. difference in cognitive load	30			[77]
	Usability	▲	Sig. better usable	31	20	25;26;27;28	[38]
	Usability	▶	No sig. difference in usability	32	25;26;32	28;29;31	[92]
Perception	Accuracy	▲	Participants were better in estimating size in larger scales in VR	33	34;35;36	99	[101]
	Accuracy	▲	Participants were better in estimating size in smaller scales in VR	34	33;35;36	99	[101]
	Accuracy	▲	Less error in height estimation in VR	35	33;34;36	99	[101]
	Accuracy	▲	Sig. lower error rate for shape and distance estimation	36	33;34;34	99	[99]
	Color	🟨	Higher luminance and chroma perception in VR	37			[102]
	Color	🟨	Higher amount of retinal illuminance in VR	38			[102]
	Efficiency	▲	Sig. higher felt individual performance in VR	39	40;42		[42]
	Efficiency	▲	VR improves perceived collaborative success	40	39;42		[95]
	Efficiency	▲	Sig. better perceived content organization	41	77		[71]
	Efficiency	🟨	Participants reported subjectively that they performed best in rich VR environment while they actually were not	42	39;40		[101]
	Engagement	🟨	Spent more time on the storytelling process when using VR	43			[56]
	Engagement	▲	Sig. higher engagement in VR	44	54		[69]
	Engagement	▶	No difference in rapport	45			[69]
	Engagement	▶	No difference in co-presence	46			[69]
	Engagement	▶	No difference in interpersonal trust	47			[69]
	Engagement	▲	VR was considered more engaging	48			[103]
	Engagement	▲	Sig. more interest and enjoyment	49	44;54		[38]
	Experience	▲	Higher immersion in VR	50	55;56;57	58	[45]
	Frustration	▲	Lower frustration levels in VR	51			[40]
	Frustration	▲	Sig. higher in perceived enjoyment	52			[71]
	Frustration	▼	Sig. higher frustration	53		44;54	[92]
	Frustration	▲	Sig. more fun in VR	54	44; 49		[104]
	Immersion	▲	Data immersion is larger in VR	55	50;56;57	58	[40]
	Immersion	▲	More immersive experience in VR	56	50;55;57	58	[56]
	Immersion	▲	Perceptual immersion higher in VR	57	50;55;56	58	[61]
	Immersion	▼	Immersion on narrative level lower in VR	58		50;55;56;57	[61]
	Learning	▶	No differences in correct insights	59		60	[41]
	Learning	▼	Less incorrect insights through VR	60		59	[41]
	Learning	▶	No differences in hypotheses generated	61		62	[41]

Table 6. *Cont.*

	Sub-Category	Tend	Finding	No.	Corr.	Contra.	Reference
Perception	Learning	▼	Fewer deep insights from within VR	62		61	[41]
	Learning	▲	User in VR can recall more information	63		64	[47]
	Learning	▼	Learned less in VR	64		63	[105]
	Learning	▲	Sig. higher motivation in learning	65			[71]
	Perception	▶	No difference in mesh resolution preferences	66			[106]
	Presence	▶	No sig. difference in presence	67	72	68;69;70;71;73	[42]
	Presence	▲	Higher presence in VR	68	69;70;71;73	67;72	[59]
	Presence	▲	Higher presence in VR condition	69	68;70;71;73	67;72	[47]
	Presence	▲	Higher presence in VR	70	69;70;71;73	67;72	[105]
	Presence	▲	Sig. stronger sense of presence	71	69;70;71;73	67;72	[38]
	Presence	▶	No sig. difference in presence	72	67	68;69;70;71;73	[92]
	Presence	▲	Sig. higher feeling of professor talking	73	69;70;71;73	67;72	[104]
	Presence	▲	Sig. higher feeling of talking to class with others	74	69;70;71;73	67;72	[104]
	Experience	▼	VR offers lower quality of experience	75	79		[100]
	Realism	▲	Meshes were perceived sig. more realistic	76			[106]
	Satisfaction	▲	Data exploration to be more satisfying in VR	77	2;11;22;23		[41]
	Satisfaction	▲	VR the most engaging	78	49		[93]
	Sim. Sick.	▼	VR induced sig. higher simulator sickness	79	75		[63]
	Spat. Perc.	▲	Better spatial perception in VR	80	33;34;35;36	99	[107]
	Workload	▼	VR shows elevation in electrodermal activity	81			[100]
	Workload	▲	Sig. lower workload in VR	82	84;88	83	[42]
	Workload	▶	No differences in workload	83		82	[40]
	Workload	▲	VR required less effort	84	82;88	83	[93]
	Workload	▼	Higher cognitive load in VR	85	86;89		[108]
	Workload	▼	Higher cognitive load in VR	86	85;89		[105]
	Workload	▶	No sig. difference in physical performance	87			[109]
	Workload	▲	Sig. lower effort	88	82;84	83	[38]
	Workload	▼	Sig. higher mental demand in VR	89	85;86		[92]
	Workload	▲	Sig. higher concentration rate in VR	90			[104]
Sens. Rec.	Accuracy	🟨	Perceived accuracy higher despite similar results	91			[93]
	Accuracy	▶	No differences in completion accuracy	92			[46]
	Accuracy	▲	Higher classification accuracy (EEG) in VR	93			[108]
	Autonomy	▲	Higher Autonomy in VR	94			[59]
	Efficiency	▶	No sig. differences in lane change performance	95			[63]
	Locomotion	🟨	Users in VR condition walked further	96			[47]
	Overview	▲	VR improves quality of view	97			[95]
	Phys. Resp.	▶	No sig. differences regarding physiological repsonses	98	87		[63]
	Spat. Perc.	▶	No difference in distance perception between all conditions	99		33;34;35;36;80	[93]
	Spat. Perc.	▼	Sig. lower realism in VR	100			[92]

[▲]: HMD VR is advantageous; [▼]: HMD VR is disadvantageous; [▶]: No difference between conditions; [🟨]: Indecisive; [**Corr.**] and [**Contr.**] describe if the finding confirms or is contradictory to other findings.

6. Discussion and Future Directions

With this work we provide an overview of 163 findings from 56 papers concerning the current research landscape on the differences and similarities between HMD VR and the real world and screen entities. All findings are grouped into three main categories *interaction*, *perception*, and *sensing and reconstructing reality*, which are further subdivided into more elaborate subcategories to evaluate differences and similarities in more detail. The study presents a summary of the research-used questionnaires (see Section 5.3) and applied study design (see Section 5.4), population (see Section 5.2), and hard-& software setup (see Section 5.1) for studies that have been conducted in the area of virtual reality research. Researchers can build on this knowledge and design more effective and rigorous

experiments. In addition, the findings from the scoping review may indicate the extent to which cross-media validity can be assumed or needs to be questioned, so that findings in one environment may or may not be transferable to another. The review of questionnaires and hardware helps to select the most appropriate measurement for their own studies, while the summary of population characteristics can help to understand the degree of the generalizability of the results.

All findings are listed and related to other findings because, as is often the case in science, there is no single truth that can be taken for granted, but rather many different aspects that need to be considered. We have identified the following three most important findings:

- In proportion, there are *more findings that show similarities* between HMD VR × real world than there are findings that show *differences* between the HMD VR and the real world. Especially for the "interaction" category as well as for the "perception" category. Only in the "sensing and reconstructing reality" category did we find more differences than similarities. This is different for HMD VR × screen, where we collected *more findings showing differences* between the HMD VR × screen environment for the *interaction* and *perception* categories. The *sensing and reconstructing reality* category is evenly distributed;

- For both entities, there are findings that need to be considered further. For example, in HMD VR × screen, learning shows mixed results (two in favor of HMD VR, two undecided, and three against). This may indicate that typical learning scenarios cannot be transferred "as is" to HMD VR, but that content and presentation type have to be adapted to the particularities of the system in order to take advantage of the specific benefits of HMD VR. This is different for HMD VR × real world where we find two results that now show differences between the two entities that could mean easier adoption; and

- When we compare the results from HMD VR with those from the real world, we observe numerous findings reporting increased symptoms of "focus difficulty", "general discomfort", "nausea", and "headache". As technology advances, we anticipate significant improvements in the design and functionality of VR systems. We predict that these advancements will effectively mitigate these prevalent issues through improved display technology, enhanced ergonomics, which includes reduced weight, an elevated user experience, and greater customization, as well as innovative algorithmic solutions;

- With an average of 28 participants (SD: 22), the study population is rather small and predominantly male.

In addition, we see an increase in software that supports setting up and conducting user studies in HMD VR for different disciplines such as "toggle toolkit" [110],"EVE" [111], or "VREX" [112]. An overview of current toolkits can be found here [113]. Platforms such as these can not only help to create a research environment, but can also help to standardize and optimize recurring features, such as the implementation of a questionnaire within the HMD VR environment.

To provide an outlook, we emphasize that the current ability of HMD VR to elicit responses and sensations that are close or similar to real-life experiences implies that HMD VR offers applications and uses beyond the often stated "gaming" purpose. In particular, HMD VR may offer promising opportunities in fields such as medicine, psychology, and other areas related to human behavior.

Although both screens and HMDs can be categorized as *technology* it is important to note that the two entities should not be treated as interchangeable. The results have shown that in many cases the outcomes are significantly different between the two entities. This does not mean that using screens to answer research questions is not a valuable approach, but it cannot replace HMDs for the reasons shown. We argue that each purpose must be evaluated individually, and efforts have to be weighed up against each other. In most cases, it can be assumed that HMD tends to produce more similar results than screens.

At present, we are nowhere near a complete understanding of how immersive VR findings can be applied to real-world outcomes, but with this work we have taken a first

step to provide direction for interested researchers. Increased research interest combined with technical advances will provide new opportunities to support knowledge transfer between HMD VR and the real world, and also to add value to established research practices, especially in cases where:

- (Attention) control is important (e.g., phobia therapy or learning situations);
- Participants are exposed to dangerous situations (e.g., firefighter training);
- Replication and sharing is useful (applies to almost any discipline except sensitive data such as patient information);
- Processes are difficult or impossible to perform in the real world (e.g., taking participants "back in time" as in reminiscence therapy); and
- Cost-efficiency is desired (e.g., participants could be recruited from anywhere in the world as long as they own an HMD).

We are confident that with recent and upcoming advances in the technology, combined with a good understanding of it, the use of immersive VR will grow for various fields that require the application of virtual studies, training, etc., to the real world. With this work, we provide a first step towards establishing a guide for a better understanding of the technology in relation to established environments, so that the respective advantages and disadvantages can be understood and implemented accordingly.

Author Contributions: Conceptualization, D.H. and M.W.; methodology, D.H.; software, D.H.; validation, D.H. and M.W.; formal analysis, D.H.; investigation, D.H.; resources, D.H.; data curation, D.H.; writing—original draft preparation, D.H.; writing—review and editing, D.H. and M.W.; visualization, D.H.; supervision, D.H. and M.W. All authors have read and agreed to the published version of the manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

CAVE	Cave automatic virtual environments
HMD	Head-mounted display
IVT	Intervention types
PRISMA	Preferred reporting items for systematic reviews and meta–analyses
VR	Virtual reality

References

1. Zheng, J.; Chan, K.; Gibson, I. Virtual reality. *IEEE Potentials* **1998**, *17*, 20–23. [CrossRef]
2. Wann, J.P.; Rushton, S.; Mon-Williams, M. Natural problems for stereoscopic depth perception in virtual environments *Vis. Res.* **1995**. *35*, 2731–2736. [CrossRef]
3. Fossataro, C.; Rossi Sebastiano, A.; Tieri, G.; Poles, K.; Galigani, M.; Pyasik, M.; Bruno, V.; Bertoni, T.; Garbarini, F. Immersive virtual reality reveals that visuo-proprioceptive discrepancy enlarges the hand-centred peripersonal space. *Neuropsychologia* **2020**, *146*, 107540. [CrossRef] [PubMed]
4. Hepperle, D.; Ödell, H.; Wölfel, M. Differences in the Uncanny Valley between Head-Mounted Displays and Monitors. In Proceedings of the 2020 International Conference on Cyberworlds (CW), Caen, France, 29 September–1 October 2020. [CrossRef]
5. Siess, A.; Wölfel, M. User color temperature preferences in immersive virtual realities. *Comput. Graph.* **2019**, *81*, 20–31. [CrossRef]
6. Cruz-Neira, C.; Sandin, D.J.; DeFanti, T.A.; Kenyon, R.V.; Hart, J.C. The CAVE: Audio visual experience automatic virtual environment. *Commun. ACM* **1992**, *35*, 64–72. [CrossRef]
7. Biocca, F. Media and the laws of the mind (Preface). In *Being There Concepts, Effects and Measurements of User Presence in Synthetic Environments*, 5th ed.; Riva, G., Davide, F., IJsselsteijn, W., Eds.; IOS Press: Amsterdam, The Netherlands, 2003; Volume 5, pp. V–VII.
8. Lohre, R.; Bois, A.J.; Athwal, G.S.; Goel, D.P.; on behalf of the Canadian Shoulder and Elbow Society (CSES)*. Improved Complex Skill Acquisition by Immersive Virtual Reality Training: A Randomized Controlled Trial. *JBJS* **2020**, *102*, e26. [CrossRef] [PubMed]

9. Cardenas, I.S.; Letdara, C.N.; Selle, B.; Kim, J.H. ImmersiFLY: Next Generation of Immersive Pilot Training. In Proceedings of the 2017 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 14–16 December 2017; pp. 1203–1206. [CrossRef]

10. Grabowski, A.; Jankowski, J. Virtual Reality-based pilot training for underground coal miners. *Saf. Sci.* **2015**, *72*, 310–314. [CrossRef]

11. Neumann, D.L.; Moffitt, R.L.; Thomas, P.R.; Loveday, K.; Watling, D.P.; Lombard, C.L.; Antonova, S.; Tremeer, M.A. A systematic review of the application of interactive virtual reality to sport. *Virtual Real.* **2018**, *22*, 183–198. [CrossRef]

12. Bordeleau, M.; Stamenkovic, A.; Tardif, P.A.; Thomas, J. The Use of Virtual Reality in Back Pain Rehabilitation: A Systematic Review and Meta-Analysis. *J. Pain* **2021**, 23, 175–195. .: 10.1016/j.jpain.2021.08.001. [CrossRef]

13. Oberhauser, M.; Dreyer, D. A virtual reality flight simulator for human factors engineering. *Cogn. Technol. Work* **2017**, *19*, 263–277. [CrossRef]

14. Tadayon, R.; Gupta, C.; Crews, D.; McDaniel, T. Do Trait Anxiety Scores Reveal Information About Our Response to Anxious Situations? A Psycho-Physiological VR Study. In Proceedings of the 4th International Workshop on Multimedia for Personal Health and Health Care, HealthMedia '19, New York, NY, USA, 21 October 2019; pp. 16–23. [CrossRef]

15. Rizzo, A.; Cukor, J.; Gerardi, M.; Alley, S.; Reist, C.; Roy, M.; Rothbaum, B.O.; Difede, J. Virtual Reality Exposure for PTSD Due to Military Combat and Terrorist Attacks. *J. Contemp. Psychother.* **2015**, *45*, 255–264. [CrossRef]

16. Yoon, J.; Byun, E.; Chung, N.S. Comparison of Space Perception between a Real Environment and a Virtual Environment. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2000**, *44*, 515–518. [CrossRef]

17. Santos, B.S.; Dias, P.; Pimentel, A.; Baggerman, J.W.; Ferreira, C.; Silva, S.; Madeira, J. Head-mounted display versus desktop for 3D navigation in virtual reality: A user study. *Multimed. Tools Appl.* **2008**, *41*, 161–181. [CrossRef]

18. Milgram, P.; Takemura, H.; Utsumi, A.; Kishino, F. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and Telepresence Technologies*; Das, H., Ed.; SPIE: Boston, MA, USA, 1995. [CrossRef]

19. Liberatore, M.J.; Wagner, W.P. Virtual, mixed, and augmented reality: A systematic review for immersive systems research. *Virtual Real.* **2021**, *25*, 773–799. [CrossRef]

20. Peters, M.D.; Godfrey, C.M.; Khalil, H.; McInerney, P.; Parker, D.; Soares, C.B. Guidance for conducting systematic scoping reviews. *Int. J. Evid. Based Healthc.* **2015**, *13*, 141–146. [CrossRef]

21. Moher, D.; Liberati, A.; Tetzlaff, J.; and, D.G.A. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef] [PubMed]

22. Hettiarachchi, A.; Wigdor, D. Annexing Reality. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016. [CrossRef]

23. Simeone, A.L.; Velloso, E.; Gellersen, H. Substitutional Reality. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems—CHI'15, Seoul, Republic of Korea, 18–23 April 2015. [CrossRef]

24. Azmandian, M.; Hancock, M.; Benko, H.; Ofek, E.; Wilson, A.D. Haptic Retargeting. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016. [CrossRef]

25. Guo, Y.; Zhang, J.; Cai, J.; Jiang, B.; Zheng, J. CNN-based Real-time Dense Face Reconstruction with Inverse-rendered Photo-realistic Face Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1294–1307. [CrossRef] [PubMed]

26. Caserman, P.; Garcia-Agundez, A.; Konrad, R.; Göbel, S.; Steinmetz, R. Real-time body tracking in virtual reality using a Vive tracker. *Virtual Real.* **2018**, *23*, 155–168. [CrossRef]

27. Rus-Calafell, M.; Garety, P.; Sason, E.; Craig, T.J.K.; Valmaggia, L.R. Virtual reality in the assessment and treatment of psychosis: A systematic review of its utility, acceptability and effectiveness. *Psychol. Med.* **2017**, *48*, 362–391. [CrossRef]

28. Munn, Z.; Peters, M.D.J.; Stern, C.; Tufanaru, C.; McArthur, A.; Aromataris, E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* **2018**, *18*, 143. [CrossRef]

29. Rosenthal, R. The file drawer problem and tolerance for null results. *Psychol. Bull.* **1979**, *86*, 638–641. [CrossRef]

30. Freitas, V. Parsifal. 2020. Available online: https://github.com/vitorfs/parsifal (accessed on 14 April 2023).

31. Gusenbauer, M.; Haddaway, N.R. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Res. Synth. Methods* **2019**, *11*, 181–217. [CrossRef] [PubMed]

32. Cole, C.L.; Kanter, A.S.; Cummens, M.; Vostinar, S.; Naeymi-Rad, F. Using a Terminology Server and Consumer Search Phrases to Help Patients Find Physicians with Particular Expertise. *Stud. Health Technol. Inform.* **2004**, *107*, 492–496. [CrossRef]

33. James, P. 3 Years Ago the Oculus Rift DK1 Shipped, Here's a Quick Look Back. 2017. Available online: https://www.roadtovr.com/3-years-ago-the-oculus-rift-dk1-shipped-heres-a-quick-look-back/ (accessed on 14 April 2023).

34. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Hum. Ment. Workload* **1988**, *1*, 139–183.

35. Witmer, B.G.; Singer, M.J. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence* **1998**, *7*, 225–240. [CrossRef]

36. Brooke, J. SUS: A "quick and dirty" usability scale. In *Usability Evaluation in Industry*; Jordan, P.W., Thomas, B., McClelland, I.L., Weerdmeester, B., Eds.; Taylor and Francis Group: Oxfordshire, UK, 1986.

37. Auer, S.; Gerken, J.; Reiterer, H.; Jetter, H.C. Comparison Between Virtual Reality and Physical Flight Simulators for Cockpit Familiarization. In Proceedings of the Mensch und Computer 2021, Ingolstadt, Germany, 5–8 September 2021. [CrossRef]

38. Elor, A.; Thang, T.; Hughes, B.P.; Crosby, A.; Phung, A.; Gonzalez, E.; Katija, K.; Haddock, S.H.D.; Martin, E.J.; Erwin, B.E.; et al. Catching Jellies in Immersive Virtual Reality: A Comparative Teleoperation Study of ROVs in Underwater Capture Tasks. In Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology, Osaka, Japan, 8–10 December 2021. [CrossRef]

39. Clifford, R.M.S.; McKenzie, T.; Lukosch, S.; Lindeman, R.W.; Hoermann, S. The Effects of Multi-sensory Aerial Firefighting Training in Virtual Reality on Situational Awareness, Workload, and Presence. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22–26 March 2020. [CrossRef]

40. Broucke, S.V.; Deligiannis, N. Visualization of Real-Time Heterogeneous Smart City Data Using Virtual Reality. In Proceedings of the 2019 IEEE International Smart Cities Conference (ISC2), Casablanca, Morocco, 14–17 October 2019; pp. 685–690.

41. Millais, P.; Jones, S.L.; Kelly, R. Exploring Data in Virtual Reality. In Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, 2018. [CrossRef]

42. Narasimha, S.; Scharett, E.; Madathil, K.C.; Bertrand, J. WeRSort: Preliminary Results from a New Method of Remote Collaboration Facilitated by Fully Immersive Virtual Reality. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2018**, *62*, 2084–2088.

43. Cai, S.; Ke, P.; Narumi, T.; Zhu, K. ThermAirGlove: A Pneumatic Glove for Thermal Perception and Material Identification in Virtual Reality. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, USA, 22–26 March 2020. [CrossRef]

44. Bahceci, O.C.; Pena-Rios, A.; Gupta, V.; Conway, A.; Owusu, G. Work-in-Progress-Using Immersive Virtual Reality in Field Service Telecom Engineers Training. In Proceedings of the 2021 7th International Conference of the Immersive Learning Research Network, Eureka, CA, USA, 17 May–10 June 2021. [CrossRef]

45. Li, Z.; Wang, J.; Yan, Z.; Wang, X.; Anwar, M.S. An Interactive Virtual Training System for Assembly and Disassembly Based on Precedence Constraints. In *Advances in Computer Graphics*; Springer Nature Publishing: Cham, Switzerland 2019; pp. 81–93. [CrossRef]

46. Pece, F.; Tompkin, J.; Pfister, H.; Kautz, J.; Theobalt, C. Device effect on panoramic video+context tasks. In Proceedings of the 11th European Conference on Visual Media Production, London, UK, 13–14 November 2014. [CrossRef]

47. Harman, J.; Brown, R.; Johnson, D. Improved Memory Elicitation in Virtual Reality: New Experimental Results and Insights. In *Human-Computer Interaction-INTERACT 2017*; Springer Nature Publishing: Cham, Switzerland 2017; pp. 128–146. [CrossRef]

48. Laugwitz, B.; Held, T.; Schrepp, M. Construction and Evaluation of a User Experience Questionnaire. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 63–76. [CrossRef]

49. Schölkopf, L.; Lorenz, M.; Stamer, M.; Albrecht, L.; Klimant, P.; Hammer, N.; Tümler, J. Haptic feedback is more important than VR experience for the user experience assessment of in-car human machine interfaces. *Procedia CIRP* **2021**, *100*, 601–606. [CrossRef]

50. Pettersson, I.; Karlsson, M.; Ghiurau, F.T. Virtually the Same Experience? In Proceedings of the 2019 on Designing Interactive Systems Conference, San Diego, CA, USA, 23–28 June 2019. [CrossRef]

51. Lewis, J.R. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *Int. J. Hum. Comput. Interact.* **1995**, *7*, 57–78. [CrossRef]

52. Schubert, T.; Friedmann, F.; Regenbrecht, H. The Experience of Presence: Factor Analytic Insights. *Presence Teleoperators Virtual Environ.* **2001**, *10*, 266–281. [CrossRef]

53. Pinto, D.; Peixoto, B.; Krassmann, A.; Melo, M.; Cabral, L.; Bessa, M. Virtual Reality in Education: Learning a Foreign Language. In *Advances in Intelligent Systems and Computing*; Springer Nature Publishing: Cham, Switzerland, 2019; pp. 589–597. [CrossRef]

54. Lewis, J.R. An after-scenario questionnaire for usability studies. *ACM SIGCHI Bull.* **1991**, *23*, 79. [CrossRef]

55. Jennett, C.; Cox, A.L.; Cairns, P.; Dhoparee, S.; Epps, A.; Tijs, T.; Walton, A. Measuring and defining the experience of immersion in games. *Int. J. Hum. Comput. Stud.* **2008**, *66*, 641–661. [CrossRef]

56. Liang, H.; Chang, J.; Deng, S.; Chen, C.; Tong, R.; Zhang, J.J. Exploitation of multiplayer interaction and development of virtual puppetry storytelling using gesture control and stereoscopic devices. *Comput. Animat. Virtual Worlds* **2016**, *28*, e1727. [CrossRef]

57. Lessiter, J.; Freeman, J.; Keogh, E.; Davidoff, J. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence Teleoperators Virtual Environ.* **2001**, *10*, 282–297. [CrossRef]

58. Ryan, R.M.; Rigby, C.S.; Przybylski, A. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motiv. Emot.* **2006**, *30*, 344–360. [CrossRef]

59. Perrin, A.; Ebrahimi, T.; Zadtootaghaj, S.; Schmidt, S.; Müller, S. Towards the need satisfaction in gaming: A comparison of different gaming platforms. In Proceedings of the 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Germany, 31 May–2 June 2017; pp. 1–3.

60. Bradley, M.M.; Lang, P.J. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **1994**, *25*, 49–59. [CrossRef]

61. Marques, T.; Vairinhos, M.; Almeida, P. How VR 360° Impacts the Immersion of the Viewer of Suspense AV Content. In Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video, TVX '19, New York, NY, USA, 5–7 June 2019; pp. 239–246. [CrossRef]

62. Kennedy, R.S.; Lane, N.E.; Berbaum, K.S.; Lilienthal, M.G. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *Int. J. Aviat. Psychol.* **1993**, *3*, 203–220. [CrossRef]

63. Weidner, F.; Hoesch, A.; Poeschl, S.; Broll, W. Comparing VR and non-VR driving simulations: An experimental user study. In Proceedings of the 2017 IEEE Virtual Reality (VR), Los Angeles, CA, USA, 18–22 March 2017; pp. 281–282. [CrossRef]

64. Lombard, M.; Ditton, T.; Weinstein, L. Measuring Presence: The Temple Presence Inventory. Presented at the Twelfth International Workshop on Presence, Los Angeles, California, USA. 2009. Available online: http://matthewlombard.com/ISPR/Proceedings/2009/Lombard_et_al.pdf (accessed on 14 April 2023).

65. Bishop, C.; Esteves, A.; McGregor, I. Head-mounted displays as opera glasses: Using mixed-reality to deliver an egalitarian user experience during live events. In Proceedings of the 19th ACM International Conference on Multimodal Interaction—ICMI 2017, Glasgow, UK, 13–17 November 2017. [CrossRef]

66. Markland, D.; Hardy, L. On the Factorial and Construct Validity of the Intrinsic Motivation Inventory: Conceptual and Operational Concerns. *Res. Q. Exerc. Sport* **1997**, *68*, 20–32. [CrossRef] [PubMed]

67. Kim, H.K.; Park, J.; Choi, Y.; Choe, M. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* **2018**, *69*, 66–73. [CrossRef]

68. O'Brien, H.L.; Toms, E.G. The development and evaluation of a survey to measure user engagement. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *61*, 50–69. [CrossRef]

69. Sanaei, M.; Machacek, M.; Eubanks, J.C.; Wu, P.; Oliver, J.; Gilbert, S.B. The Effect of Training Communication Medium on the Social Constructs Co-Presence, Engagement, Rapport, and Trust. In Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology, Tsukuba, Japan, 29 November–1 December 2022. [CrossRef]

70. Franklin, A.E.; Burns, P.; Lee, C.S. Psychometric testing on the NLN Student Satisfaction and Self-Confidence in Learning, Simulation Design Scale, and Educational Practices Questionnaire using a sample of pre-licensure novice nurses. *Nurse Educ. Today* **2014**, *34*, 1298–1304. [CrossRef]

71. Hoang, T.; Greuter, S.; Taylor, S. An Evaluation of Virtual Reality Maintenance Training for Industrial Hydraulic Machines. In Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Christchurch, New Zealand, 12–16 March 2022. [CrossRef]

72. Lanier, M.; Waddell, T.; Elson, M.; Tamul, D.; Ivory, J.; Przybylski, A. Virtual reality check: Statistical power, reported results, and the validity of research on the psychology of virtual reality and immersive environments. *Comput. Hum. Behav.* **2019**, *100*, 70–78. [CrossRef]

73. Takahashi, N.; Inamura, T.; Mizuchi, Y.; Choi, Y. Evaluation of the Difference of Human Behavior between VR and Real Environments in Searching and Manipulating Objects in a Domestic Environment. In Proceedings of the 2021 30th IEEE International Conference on Robot Human Interactive Communication, Vancouver, BC, Canada, 8–12 August 2021. [CrossRef]

74. Zavlanou, C.; Lanitis, A. Product Packaging Evaluation Through the Eyes of Elderly People: Personas vs. Aging Suit vs. Virtual Reality Aging Simulation. In *Human Systems Engineering and Design*; Springer Nature Switzerland AG: Cham, Switzerland, 2018; pp. 567–572. [CrossRef]

75. Han, H.; Lu, A.; Wells, U. Under the Movement of Head: Evaluating Visual Attention in Immersive Virtual Reality Environment. In Proceedings of the 2017 International Conference on Virtual Reality and Visualization (ICVRV), Zhengzhou, China, 21–22 October 2017; pp. 294–295.

76. Ebrahimi, E.; Babu, S.V.; Pagano, C.C.; Jörg, S. An Empirical Evaluation of Visuo-Haptic Feedback on Physical Reaching Behaviors During 3D Interaction in Real and Immersive Virtual Environments. *ACM Trans. Appl. Percept.* **2016**, *13*, 1–21. [CrossRef]

77. Mathur, J.; Miller, S.R.; Simpson, T.W.; Meisel, N.A. Identifying the Effects of Immersion on Design for Additive Manufacturing Evaluation of Designs of Varying Manufacturability. In Proceedings of the Volume 5: 27th Design for Manufacturing and the Life Cycle Conference (DFMLC), St. Louis, MO, USA, 14–17 August 2022. [CrossRef]

78. Mathis, F.; Vaniea, K.; Khamis, M. RepliCueAuth: Validating the Use of a Lab-Based Virtual Reality Setup for Evaluating Authentication Systems. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021. [CrossRef]

79. Ma, C.; Han, T. Combining Virtual Reality (VR) Technology with Physical Models – A New Way for Human-Vehicle Interaction Simulation and Usability Evaluation. In *HCI in Mobility, Transport, and Automotive Systems*; Krömker, H., Ed.; Springer Nature Switzerland AG: Cham, Switzerland, 2019; pp. 145–160.

80. Chew, J.Y.; Okayama, K.; Okuma, T.; Kawamoto, M.; Onda, H.; Kato, N. Development of A Virtual Environment to Realize Human-Machine Interaction of Forklift Operation. In Proceedings of the 2019 7th International Conference on Robot Intelligence Technology and Applications (RiTA), Daejeon, Republic of Korea, 1–3 November 2019; pp. 112–118.

81. Verwulgen, S.; Goethem, S.V.; Cornelis, G.; Verlinden, J.; Coppens, T. Appreciation of Proportion in Architecture: A Comparison Between Facades Primed in Virtual Reality and on Paper. In *Advances in Human Factors in Wearable Technologies and Game Design*; Springer Nature Publishing: Cham, Switzerland, 2019; pp. 305–314. [CrossRef]

82. Liao, D.; Zhang, W.; Liang, G.; Li, Y.; Xie, J.; Zhu, L.; Xu, X.; Shu, L. Arousal Evaluation of VR Affective Scenes Based on HR and SAM. In Proceedings of the 2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC), Nanjing, China, 6–8 May 2019; Volume 1, pp. 1–4.

83. Ostrander, J.K.; Tucker, C.S.; Simpson, T.W.; Meisel, N.A. Evaluating the Effectiveness of Virtual Reality As an Interactive Educational Resource for Additive Manufacturing. In Proceedings of the Volume 3: 20th International Conference on Advanced Vehicle Technologies 15th International Conference on Design Education, Quebec City, QC, Canada, 26–29 August 2018. [CrossRef]

84. Keller, T.; Brucker-Kley, E.; Wyder, C. Virtual reality and its impact on learning success. In Proceedings of the 16th International Conference Mobile Learning 2020, Sofia, Bulgaria, 2–4 April 2020; pp. 78–86.

85. Guo, J.; Weng, D.; Fang, H.; Zhang, Z.; Ping, J.; Liu, Y.; Wang, Y. Exploring the Differences of Visual Discomfort Caused by Long-term Immersion between Virtual Environments and Physical Environments. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, USA, 22–26 March 2020. [CrossRef]

86. Diederichs, F.; Niehaus, F.; Hees, L. Guerilla Evaluation of Truck HMI with VR. In *Virtual, Augmented and Mixed Reality. Design and Interaction*; Springer Nature Switzerland AG: Cham, Switzerland, 2020; pp. 3–17. [CrossRef]

87. Vazquez, C.; Xia, L.; Aikawa, T.; Maes, P. Words in Motion: Kinesthetic Language Learning in Virtual Reality. In Proceedings of the 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), Mumbai, India, 9–13 July 2018. [CrossRef]

88. Agethen, P.; Link, M.; Gaisbauer, F.; Pfeiffer, T.; Rukzio, E. Counterbalancing virtual reality induced temporal disparities of human locomotion for the manufacturing industry. In Proceedings of the 11th Annual International Conference on Motion, Interaction, and Games, Limassol, Cyprus, 8–10 November 2018. [CrossRef]

89. Franzluebbers, A.; Johnsen, K. Performance Benefits of High-Fidelity Passive Haptic Feedback in Virtual Reality Training. In Proceedings of the Symposium on Spatial User Interaction—SUI'18, Berlin, Germany, 13–14 October 2018. [CrossRef]

90. Bialkova, S.; Ettema, D. Cycling renaissance: The VR potential in exploring static and moving environment elements. In Proceedings of the 2019 IEEE 5th Workshop on Everyday Virtual Reality (WEVR), Osaka, Japan, 23–24 March 2019; pp. 1–6.

91. Christensen, D.J.R.; Holte, M.B. The Impact of Virtual Reality Training on Patient-Therapist Interaction. In *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*; Springer International Publishing AG: Cham, Switzerland, 2018; pp. 127–138. . [CrossRef]

92. Safikhani, S.; Holly, M.; Kainz, A.; Pirker, J. The Influence of in-VR Questionnaire Design on the User Experience. In Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology, Osaka, Japan, 8–10 December 2021. [CrossRef]

93. Wagner Filho, J.A.; Rey, M.F.; Freitas, C.M.D.S.; Nedel, L. Immersive Visualization of Abstract Information: An Evaluation on Dimensionally-Reduced Data Scatterplots. In Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Tuebingen/Reutlingen, Germany, 18–22 March 2018; pp. 483–490.

94. Petrykowski, M.; Berger, P.; Hennig, P.; Meinel, C. Digital Collaboration with a Whiteboard in Virtual Reality. In *Proceedings of the Future Technologies Conference (FTC) 2018*; Springer Nature Switzerland AG: Cham, Switzerland, 2018; pp. 962–981. [CrossRef]

95. Kratz, S.; Rabelo Ferriera, F. Immersed remotely: Evaluating the use of Head Mounted Devices for remote collaboration in robotic telepresence. In Proceedings of the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016; pp. 638–645.

96. Andersen, B.J.H.; Davis, A.T.A.; Weber, G.; Wunsche, B.C. Immersion or Diversion: Does Virtual Reality Make Data Visualisation More Effective? In Proceedings of the 2019 International Conference on Electronics, Information, and Communication (ICEIC), Auckland, New Zealand, 22–25 January 2019. [CrossRef]

97. Franzluebbers, A.; Li, C.; Paterson, A.; Johnsen, K. Virtual Reality Point Cloud Annotation. In Proceedings of the 2022 ACM Symposium on Spatial User Interaction, Online, 1–2 December 2022. [CrossRef]

98. Zhang, X.; He, W.; Wang, S. Manual Preliminary Coarse Alignment of 3D Point Clouds in Virtual Reality. In *Communications in Computer and Information Science*; Springer Nature Switzerland AG, Cham, Switzerland 2021; pp. 424–432. [CrossRef]

99. Hombeck, J.; Meuschke, M.; Zyla, L.; Heuser, A.J.; Toader, J.; Popp, F.; Bruns, C.J.; Hansen, C.; Datta, R.R.; Lawonn, K. Evaluating Perceptional Tasks for Medicine: A Comparative User Study Between a Virtual Reality and a Desktop Application. In Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Christchurch, New Zealand, 12–16 March 2022. [CrossRef]

100. Keighrey, C.; Flynn, R.; Murray, S.; Murray, N. A Physiology-based QoE Comparison of Interactive Augmented Reality, Virtual Reality and Tablet-based Applications. *IEEE Trans. Multimed.* **2020**, *23*, 333–341. [CrossRef]

101. Watson, D.; Fitzmaurice, G.; Matejka, J. How Tall is that Bar Chart? Virtual Reality, Distance Compression and Visualizations. In Proceedings of the 2021 Graphics Interface Conference, Virtual Event, 28–29 May 2021; Canadian Information Processing Society: Mississauga, ON, Canada, 2021. [CrossRef]

102. Nishimura, T.; Hirai, K.; Horiuchi, T. Color Perception Comparison of Scene Images between Head-Mounted Display and Desktop Display. In Proceedings of the International Display Workshops, Sapporo, Japan, 27–29 November 2019; p. 1148. [CrossRef]

103. Nebeling, M.; Rajaram, S.; Wu, L.; Cheng, Y.; Herskovitz, J. XRStudio: A Virtual Production and Live Streaming System for Immersive Instructional Experiences. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021. [CrossRef]

104. Fujii, R.; Hirose, H.; Aoyagi, S.; Yamamoto, M. On-Demand Lectures that Enable Students to Feel the Sense of a Classroom with Students Who Learn Together. In *Human Interface and the Management of Information. Information Presentation and Visualization*; Springer Nature Switzerland AG: Cham, Switzerland, 2021; pp. 268–282. [CrossRef]

105. Makransky, G.; Terkildsen, T.S.; Mayer, R.E. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn. Instr.* **2019**, *60*, 225–236. [CrossRef]

106. Thorn, J.; Pizarro, R.; Spanlang, B.; Bermell-Garcia, P.; Gonzalez-Franco, M. Assessing 3D Scan Quality Through Paired-comparisons Psychophysics. In Proceedings of the 2016 ACM on Multimedia Conference—MM'16, Amsterdam, The Netherlands, 15–19 October 2016. [CrossRef]

107. Horvat, N.; Škec, S.; Martinec, T.; Lukacevic, F.; Perišic, M. Comparing virtual reality and desktop interface for reviewing 3D CAD models. In Proceedings of the Design Society: International Conference on Engineering Design, Delft, The Netherlands, 5–8 August 2019; pp. 1923–1932. [CrossRef]

108. Qadir, Z.; Chowdhury, E.; Ghosh, L.; Konar, A. Quantitative Analysis of Cognitive Load Test While Driving in a VR vs. Non-VR Environment. In *Lecture Notes in Computer Science*; Springer Nature Switzerland AG: Cham, Switzerland, 2019; pp. 481–489. [CrossRef]

109. Colombo, V.; Bocca, G.; Mondellini, M.; Sacco, M.; Aliverti, A. Evaluating the effects of Virtual Reality on perceived effort during cycling: Preliminary results on healthy young adults. In Proceedings of the 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Messina, Italy, 22–24 June 2022. [CrossRef]

110. Ugwitz, P.; Šašinková, A.; Šašinka, v.; Stachoň, Z.; Juřík, V. Toggle toolkit: A tool for conducting experiments in unity virtual environments. *Behav. Res. Methods* **2021**, *53*, 1581–1591. [CrossRef]

111. Grübel, J.; Weibel, R.; Jiang, M.H.; Hölscher, C.; Hackman, D.A.; Schinazi, V.R. EVE: A framework for experiments in virtual environments. In *Spatial Cognition X*; Springer International Publishing AG: Cham, Switzerland, 2016; pp. 159–176.

112. Vasser, M.; Kängsepp, M.; Magomedkerimov, M.; Kilvits, K.; Stafinjak, V.; Kivisik, T.; Vicente, R.; Aru, J. VREX: An open-source toolbox for creating 3D virtual reality experiments. *BMC Psychol.* **2017**, *5*, 4. [CrossRef] [PubMed]

113. Wolfel, M.; Hepperle, D.; Purps, C.F.; Deuchler, J.; Hettmann, W. Entering a new Dimension in Virtual Reality Research: An Overview of Existing Toolkits, their Features and Challenges. In Proceedings of the 2021 International Conference on Cyberworlds (CW), Caen, France, 28–30 September 2021. [CrossRef]