A Critical Assessment of the Use of SSQ as a Measure of General Discomfort in VR Head-Mounted Displays

Teresa Hirzle
Institute of Media Informatics, Ulm University
Ulm, Germany
teresa.hirzle@uni-ulm.de

Maurice Cordts
Institute of Media Informatics, Ulm University
Ulm, Germany
maurice.cordts@uni-ulm.de

Enrico Rukzio
Institute of Media Informatics, Ulm University
Ulm, Germany
enrico.rukzio@uni-ulm.de

Jan Gugenheimer
Télécom Paris - LTCI, Institut Polytechnique de Paris
Paris, France
jan.gugenheimer@telecom-paris.fr

Andreas Bulling
Institute for Visualisation and Interactive Systems, University of Stuttgart
Stuttgart, Germany
andreas.bulling@vis.uni-stuttgart.de

ABSTRACT

Based on a systematic literature review of more than 300 papers published over the last 10 years, we provide indicators that the simulator sickness questionnaire (SSQ) is extensively used and widely accepted as a general discomfort measure in virtual reality (VR) research – although it actually only accounts for one category of symptoms. This results in important other categories (digital eye strain (DES) and ergonomics) being largely neglected. To contribute to a more comprehensive picture of discomfort in VR head-mounted displays, we further conducted an online study (N=352) on the severity and relevance of all three symptom categories. Most importantly, our results reveal that symptoms of simulator sickness are significantly less severe and of lower prevalence than those of DES and ergonomics. In light of these findings, we critically discuss the current use of SSQ as the only discomfort measure and propose a more comprehensive factor model that also includes DES and ergonomics.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **User studies**; **Virtual reality**.

KEYWORDS

SSQ, simulator sickness, digital eye strain, discomfort, head-mounted displays, virtual reality

ACM Reference Format:
Teresa Hirzle, Maurice Cordts, Enrico Rukzio, Jan Gugenheimer, and Andreas Bulling. 2021. A Critical Assessment of the Use of SSQ as a Measure of General Discomfort in VR Head-Mounted Displays. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3411764.3445361

CHI '21, May 8–13, 2021, Yokohama, Japan
2021. ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445361

1 INTRODUCTION

Feelings of discomfort or even sickness are well-known, frequent, and often unavoidable byproducts when using VR head-mounted displays (HMDs) [93]. Several reviews on discomfort in VR have investigated and presented possible causes for discomfort, including hardware, content, as well as causes related to human factors [18, 80, 84]. Others have demonstrated that effects of discomfort in VR HMDs – also known as VR sickness [18, 49], cybersickness [22, 80, 85], or visually induced motion sickness [45] – are distinct from those of simulator sickness [62, 88], which is typically measured with the simulator sickness questionnaire (SSQ) [46]. The SSQ was introduced in the 90s for the assessment of sickness symptoms during professional flight simulator training [18, 84]. While revised questionnaires specifically for VR settings based on the SSQ were proposed, such as the Virtual Reality Sickness Questionnaire [49] or the Revised SSQ [48], these only provide a modified factor structure and do not include other symptom categories of discomfort. Single item measures asking participants to rate their sickness on a linear scale are commonly used alternatives but further reduce the complexity of discomfort [25, 47, 56].

When measured with the SSQ, symptoms in VR result in higher and a different distribution of severity scores across the three sub scales (nausea, oculomotor, disorientation) than simulator sickness, with disorientation scoring highest, followed by nausea, and oculomotor symptoms. In contrast, in simulator sickness studies, oculomotor symptoms scored highest, followed by nausea and disorientation [84, 88]. Moreover, sickness in simulators and VR HMDs has been understood as a form of motion sickness that is caused by conflicting visual and vestibular information [62]. Therefore, given that it was developed based on the Pensacola Motion Sickness Questionnaire (MSQ) [43], the SSQ mainly addresses symptoms of motion sickness. However, in contrast to simulator sickness, symptoms in VR can occur with visual stimulation and the absence of vestibular stimulation [62], which causes an illusory feeling of self-motion called vection [33]. Finally, and most importantly, symptoms of DES [3, 82] or symptoms concerning the headset's ergonomics (ERG) [28] were found being important contributors to discomfort

in VR HMDs, which indicates that discomfort symptomatology in VR HMDs includes more than motion sickness symptoms.

This work makes three original contributions to address the aforementioned limitations and, thus, contribute to a better understanding of discomfort in VR HMDs. First, we conducted a systematic literature review on the current use of SSQ in VR studies. Our review covers more than 300 papers published over the last 10 years in three academic databases (ACM DL, IEEE Xplore, and ScienceDirect). We found that the SSQ is indeed being used as a tool to assess general discomfort in HMDs rather than being specifically used to assess simulator sickness. In addition, important additional factors of discomfort (DES and ERG) are largely neglected in VR research. To evaluate the relationship of simulator sickness, DES, and ERG as well as their importance for current VR users, we then conducted a large-scale online user study (N=352). Results from our study show that symptoms of simulator sickness occur less severely and are less relevant to users actually owning and using the technology in comparison to DES and ERG symptoms. Using a series of factor analyses, we further suggest that the factor structure of discomfort addressing these three categories in VR HMDs is comprised by six factors with ERG symptoms contributing most to discomfort, followed by DES and simulator sickness. The specific contributions of this work are:

- A systematic literature review covering more than 300 papers that shows that the SSQ has become the de facto standard without necessarily questioning whether it fits the research question. And despite the fact that SSQ was (1) developed for a different purpose and (2) only covers a single category of symptoms of discomfort.
- An extensive online user study (N=352) showing that simulator sickness symptoms are less severe and less important to frequent VR HMD users than symptoms of ERG and DES.
- A more comprehensive factor model of discomfort in VR HMDs, suggesting that discomfort in VR HMDs is comprised by (at least) three orthogonal, independent factors (ERG, DES, and simulator sickness).

We hope that these findings will trigger a discussion and rethinking in the community of using SSQ as the main measure of discomfort. Our analysis reveals that it is important to shift focus towards more types of symptoms (such as ERG and DES symptoms) that affect users of VR headsets right now and in future. We believe that our results have the potential to lay the foundation for the development of a more sensitive and comprehensive measure in the future.

2 SYSTEMATIC LITERATURE REVIEW ON CURRENT PRACTICE IN THE USE OF SSQ

A critical assessment of SSQ as a measure of discomfort in VR HMDs first requires to understand why and how the questionnaire is currently being used in the field and how its results are being reported in research papers. To this end, we conducted a systematic literature review with meta-analysis covering more than 300 papers published over the last 10 years. In stark contrast to other recent reviews of SSQ, we intentionally did not focus on influencing factors or causes of simulator sickness [18, 80, 84], but instead on the authors' rationales for using it. To conduct the review, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [65] that were previously used to review factors causing simulator sickness in VR HMDs [84].

2.1 Method

2.1.1 Identification of Sources. We searched in the following three academic databases: ACM Digital Library (ACM DL)[1], IEEE Xplore[2], and Science Direct[3]. We opted for these databases given that they cover the most relevant conferences and journals for human-computer interaction and VR research (e.g., CHI, IEEE VR, IEEE ISMAR). The search was conducted between July 1, 2020 (ACM DL and IEEE Xplore) and July 22, 2020 (ScienceDirect). We searched the databases' full text collections with the combined term "simulator sickness questionnaire" between the years 2010 and 2020. This initial search resulted in 833 papers. The summary of the following screening process is shown in Figure 1.

2.1.2 Screening of Relevant Research. In the first screening phase, two of the authors read through the title, abstract, and reference sections of all 833 papers. 247 papers were excluded based on the following *exclusion criteria*: no access to full text (115, e.g., Ling et al. [66]), no citation of SSQ (67, e.g., when a simulator sickness questionnaire was mentioned but not cited, such as Kaber et al. [42]), meta papers (34, e.g., surveys and reviews, such as Grubert et al. [31]), not written in English (12, e.g., Gonçalves et al. [29]), duplicates (12, i.e., the same study or paper at a different venue with the same or different title, in this case only the earliest publication was included e.g., Kutsuna et al. [92] (excluded), Kutsuna et al. [59] (kept)), and not part of the conference proceedings (7, e.g., a doctoral consortium, such as Maloney et al. [71]).

2.1.3 Eligibility. In the second screening phase, the same two authors assessed the eligibility of the remaining 586 papers based on the following *inclusion criteria*. Only papers were included that presented a study in which the SSQ was used and in which mean values were reported. 189 papers were excluded that stated to have used the SSQ, but did not report on the results (e.g., D'Angelo et al. [21]).

Another 57 papers were excluded, because they did not report absolute, but only relative values (e.g., Polonen et al. [79]), 12 did not actually employ the SSQ (e.g. da Costa et al. [60]), in 12 the data was not extractable (e.g., because values were reported on a different scale and the conversion was not transparently reported, such as Lugrin et al. [69]), and 7 used a modified version of SSQ (e.g., Lopez et al [67]).

2.1.4 Analysis of Included Papers. Papers were annotated with regard to the *rationale* for employing the SSQ, the mean values for the *total score*, as well as for *sub scale scores* and *single item values*. We further coded whether only *post-* or *pre- and post - exposure* values were reported, whether a visuo-vestibular *conflict* was present, the *type of system* that was used, the *gender distribution* of the sample that was reported, and whether *additional DES or ERG measures* were employed. Initially we also coded whether papers only reported on *pre-exposure*, but since no paper of the

[1] https://dl.acm.org/, accessed September 10th, 2020
[2] https://ieeexplore.ieee.org/, accessed September 10th, 2020
[3] https://www.sciencedirect.com/, accessed September 10th, 2020

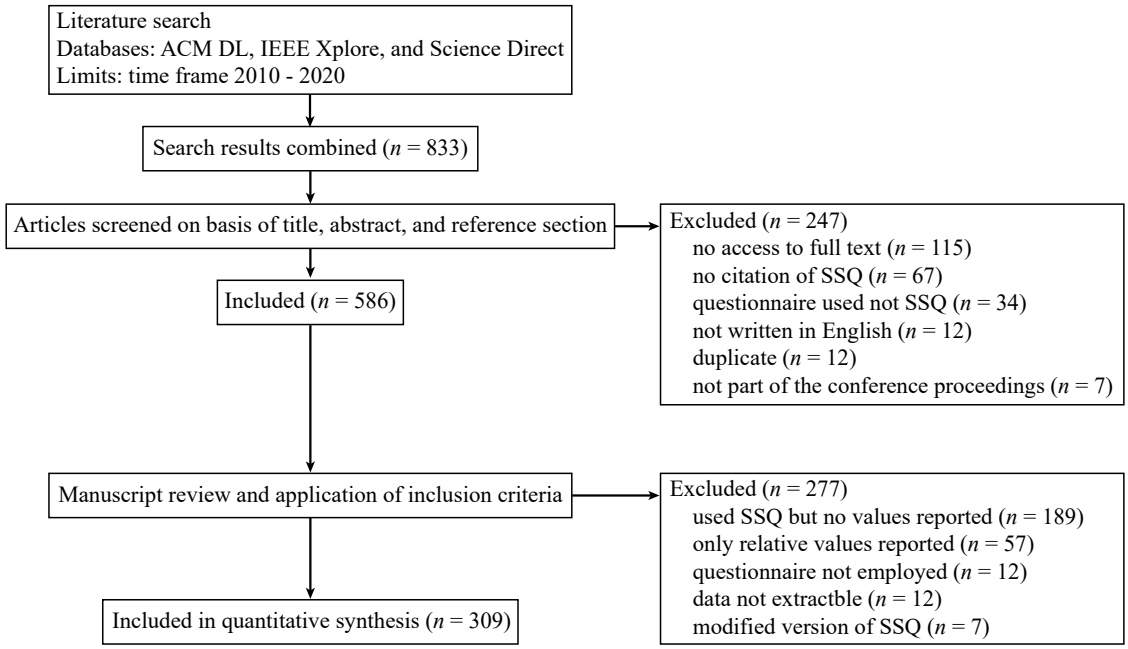

Figure 1: PRISMA flow diagram of selection process.

data set solely reported on this, we left that out. The final set of papers with all annotations and exclusion criteria is available in the supplementary material.

2.2 Data Analysis and Results

2.2.1 Rationale for Using the SSQ. In roughly half of the final papers (170/55%) authors clearly stated that they used the SSQ to assess either simulator sickness (113/37%, e.g., Abd-Alhamid [1]), cybersickness (43/14%, e.g., Benzina et al. [6]), visually induced motion sickness (10/3%, e.g., [90]), or VR sickness (5/2%, e.g., Lubeck et al. [68]). The vast majority of the remaining papers (82/27%) did not provide a reason for employing SSQ (e.g., Bolte et al. [10]). When no specific reasons were provided, we concluded that authors used the SSQ as a standard measure of sickness symptoms in VR. Of the remaining papers, 25 (8%) provided motion sickness (e.g., Chen et al. [19]) and 9 papers (3%) discomfort (e.g., Lai et al. [61]) as rationale. The remaining 17 papers (6%) provided different reasons, such as to assess health status (5), visual fatigue (4), or physiological side effects of participants (3) (e.g., Krekhov et al. [57]).

2.2.2 Visuo-vestibular Conflict. In 189 papers (61%) a visuo-vestibular conflict was present – a conflict widely known as one major cause of simulator sickness symptoms [62]. However, in 41 of 113 (34%) papers that reported to measure simulator sickness no conflict was present. Even in the 8 of 25 papers (32%) that reported to measure motion sickness, no conflict was found. We made a similar observation for papers that reported cybersickness, where in 13 out of 43 papers (30%) no conflict was present.

2.2.3 Administration and Reported Values. Of the reviewed papers, 119/39% administered the SSQ before (pre-exposure) and after (post-exposure) the experiment, while 190/61% employed it only post-exposure. None of the papers employed the SSQ solely pre-exposure. 285 of the papers (92%) reported a SSQ total score. In the remaining 8% of the papers only sub scale values were reported. The mean of the SSQ total score was 24.90. In 175 papers (57%) the SSQ was employed as a post-exposure measure with a mean total score of 23.95. In 110 papers (36%) the SSQ was employed as pre- and post-exposure measure with a mean total score of 25.85. Of all papers, 135 (44%) reported sub scale values. The mean post-exposure values for the sub scales were: nausea (23.74), oculomotor (24.17), and disorientation (28.8). The total post-exposure score of papers with visuo-vestibular conflict was slightly higher (26.29) than for papers where no conflict was present (22.08).

2.2.4 Gender, Device Type, and Additional Measures. Of all papers, 162 (52%) conducted a study with a HMD, followed by 73/24% that reported a virtual environment as apparatus. We found 27 papers (9%) that reported to have used a stereoscopic display (other than HMD), 18 that used a driving simulator (6%), 13 studies in a CAVE (4%), 9 that used a display screen (3%), and 7 that reported diverse other device types (2%). The mean gender distribution of all papers was 39% female and 61% male participants, with 4002 female, 6259 male, and 2 non-binary participants. We found only 13 papers (4%) that applied an additional measure to assess digital eye strain and none that assessed additional ergonomic symptoms – despite the fact that some reported that participants experienced symptoms from "wearing the device" [23].

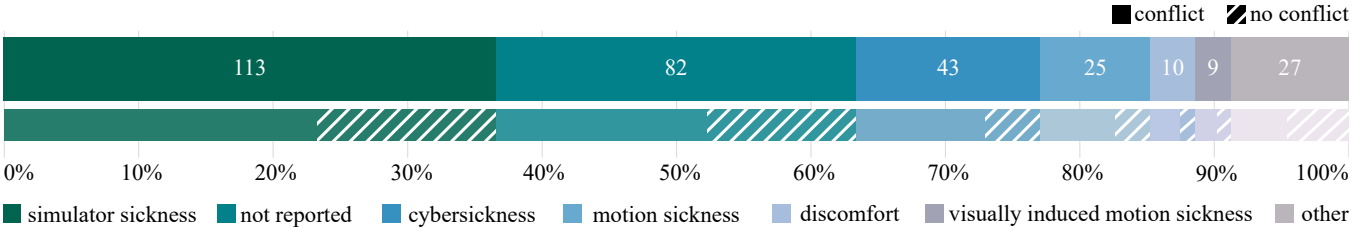

Figure 2: Reported rationale for employing SSQ.

3 SSQ AS A MEASURE OF GENERAL DISCOMFORT

As our literature review has revealed, SSQ is not only widely used in the field, and in inconsistent ways (post-exposure only vs. pre- and post-exposure), but also typically as the only measure and in many cases even under the (wrong) assumption of it as a general measure of discomfort. In addition to problems related to the common practice of using the SSQ, the questionnaire itself has been criticised for different other reasons in the past.

Correlated factor structure. The SSQ was derived from 1,119 motion sickness questionnaire samples [43] covering 16 of 28 symptoms measured in simulator studies with 10 different simulators [46]. Its three-factor model was derived by a principal factor analysis with varimax rotation. As factors were correlated, the authors conducted a hierarchical factor analysis to extract a general factor that all items had nonzero loadings on, in addition to three group factors. This resulted in the final factor structure of the SSQ, which is comprised by a total score and the three sub scale factors nausea, oculomotor, and disorientation. The factor structure of the SSQ was revised several times, mostly with the aim to propose uncorrelated, orthogonal factors. Bouchard et al. [12] proposed a two-factor, orthogonal solution for SSQ symptoms, with one factor being nausea symptoms and one oculomotor symptoms, following the same method as Kennedy et al. (principal factor analysis with varimax rotation) [46]. They were able to extract factors that had no cross-loading items with loadings below 0.4. The authors attributed the differences in factor structure to differences in sample, device, and task. However, it cannot be ruled out that the French translation of the SSQ that was used had an effect, too. Balk et al. found fairly similar symptom loadings as the SSQ with a data set of nine driving simulator studies [4]. However, five symptoms (burping, fatigue, headache, blurred vision, and fullness of head) were not attributed to any factor and they did not have items with cross-loadings. Based on their results, the authors suggested to revise the factor structure and reevaluate the factor weights of the SSQ when using it as a diagnostic tool.

Limited suitability as a measure of symptoms in VR.. Sevinc and Ilker investigated psychometric qualities (e.g., construct validity, test-retest reliability) of two modified SSQ versions (the Virtual Reality Sickness Questionnaire (VRSQ) [49] and the Cybersickness Questionnaire (CSQ) [89]), in comparison to the SSQ to test their quality as cybersickness measures [85]. Based on their data, they found construct validity for both VRSQ and CSQ, but not for the SSQ. Although they addressed fewer symptoms than SSQ, CSQ and VRSQ were more sensitive in detecting differences in different VR applications. Finally, Bruck and Watters derived a four-factor model (cybersickness, vision, arousal, and fatigue) for cybersickness based on 28 SSQ samples [13]. However, similar to SSQ, their analysis produced correlated factors. In addition, they applied principal component analysis, which in contrast to factor analysis does not detect latent variables, but finds an optimal linear combination of components [41].

Sample does not provide generalizability. The questionnaire was developed for an expert user group (pilots) using a specific training system. Today's VR HMDs are everyday devices, i.e., users use devices self-motivated and not as a prerequisite for their professional carrier. To investigate symptoms in this specific population, we conducted an online study that enabled us to have access to frequent VR HMD users. Furthermore, the SSQ was derived by 1,119 samples of male pilots. Additionally, it was reported that women are more susceptible to motion sickness [26] and report higher values of related symptoms [2, 7, 27]. This means that the SSQ, especially the sub scale multipliers, are strongly biased towards mens' perception of symptoms, which might be significantly different for women. To address these limitations in our analysis of general discomfort in VR HMDs, we analyse a sex-balanced sample and evaluate possible differences between men and women.

Limited scope of use. The SSQ was developed as a measure of simulator sickness and therefore does not allow to assess other symptoms that are of equal, if not even higher importance in the context of VR HMDs, specifically digital eye strain and ergonomic symptoms. Szpak et al. stressed visual fatigue and cognitive fatigue as additional influence factors that are not considered sufficiently by the SSQ [91]. Similarly, Ames et al. criticised that the SSQ does not cover the range of ocular symptoms that can occur in VR HMDs appropriately and designed the Virtual Reality Symptom Questionnaire to stress the importance of ocular symptoms [3]. Ocular symptoms that arise from prolonged viewing of digital devices are summarized by the term *Digital Eye Strain*, which refers to vision and eye problems that are reported by users after experiencing prolonged screen time [81, 86]. Symptoms include general discomfort, headache, fatigue, but also address specific properties of the eyes, such as tearing or burning eyes [24, 40, 86]. Numerous causes were reported, such as ocular anomalies [81], reduced blink rate [9], close viewing distances [37], or interface properties [82]. In addition to problems caused by digital screens in general, HMDs pose specific challenges to the eyes [36], for instance by a particularly short screen-eye distance, and most importantly

Table 1: The symptoms of simulator sickness and digital eye strain included in our online study.

Simulator Sickness Symptoms	Digital Eye Strain Symptoms
General discomfort, Fatigue, Headache, Eyestrain, Difficulty focusing, Increased salivation, Sweating, Nausea, Difficulty concentrating, Fullness of head, Blurred vision, Dizzyness with open eyes, Dizzyness with closed eyes, Vertigo, Stomach awareness, Burping	Burning eyes, Double vision, Dry eyes, Excessive blinking, Eye ache, Eye redness, Feeling of a foreign body, Feeling that sight is worsening, Heavy eyelids, Increased sensitivity to light, Irritated eyes, Neck pain, Seeing colored halos around objects, Sensation of hot eyes, Shoulder Pain, Soreness of eyes, Tearing eyes, Watering of eyes

the vergence-accommodation conflict [78], which is a well-known source of digital eye strain in HMDs [50, 87]. While the simulator sickness questionnaire contains an oculomotor sub scale with 7 symptoms, literature on DES indicates that a more differentiated look at the importance of eye strain for discomfort in VR HMDs is needed. HMDs are wearable devices that are attached to the body but only few studies investigated symptoms caused by the ergonomic factors of HMDs. Motti and Kelly identified several factors that are important to users when buying a headset [28]. Among the top ten factors are weight, size, and comfort of the headset. Users adopt for the weight of devices by buying head pads that mean to reduce pressure on the head and cheeks, and to redistribute the weight on the head. Not surprisingly, in 2007 Knight and Baber found that wearing a HMD causes users to significantly change their neck posture, which may add increased levels of stress to the musculoskeletal system [53]. These types of ergonomic symptoms have gained significance with VR HMDs entering users' homes. As they have only become important with the wearable form factor of simulator technology, it is comprehensible that they were not included in the SSQ.

4 EVALUATION OF ERG, DES, AND SSQ AS MEASURES OF DISCOMFORT IN VR HMDS

The widespread and often unquestioned use of SSQ as a general discomfort measure in VR HMDs – as shown by our survey and discussed in specific previous works – motivated us to study the prevalence and importance of additional symptom categories in more detail. We chose ERG and DES given that they are important and frequently occurring symptoms in closely related fields of research – when studying comfort of wearable devices [52] and with users who are exposed to prolonged screen time [9, 24]. To be able to compare the prevalence and severity of ERG, DES, and simulator sickness, we designed a between-subject online user study that we detail in the following.

4.1 Method

4.1.1 Selection of Symptoms. To assess *simulator sickness* we used the 16 items of the original SSQ (Table 1 left). The items for *digital eye strain* are based on Hirzle et al.'s review on digital eye strain in gaze-based interactive systems [35]. We finally added nine items that were additionally reported in the medical domain (Blehm et al.'s overview of symptoms [9] and the Computer Vision Syndrome-Questionnaire [24]), resulting in 18 different symptoms (Table 1 right).

To assess *ergonomic symptoms*, we used the six comfort rating scales proposed by Knight and Baber (see Table 2) [52, 54]. Six statements covering the attachment of the device were added inspired by Cancela et al.'s wearability assessment of a system for Parkinson's disease patients [15]. Further, Borg scales [11] have a long history of being employed to evaluate exertion and perceived pain of wearable devices [15, 55], including HMDs [51]. The Borg CR10 (category-ratio) scale can be used together with a body map on which participants indicate the level of pain or discomfort they are currently experiencing on different regions of their body. We therefore employed a scale based on a Borg CR10 scale to localise pain and discomfort in different regions of the head, face, and neck caused by the attachment of the headset (see Figure 4 for face map based on surface anatomy of the face and neck that was used in the study).

4.1.2 Definition of Rating Scale. To ensure comparability of the three symptom categories, we aimed for uniform rating scales. This confronted us with the challenge of integrating more than four types of scales (SSQ scale [46], several scales for DES [24, 40, 86], CRS [54], and Borg CR10 scale [11]) that varied in length and sensitivity. In general, the number of response categories is closely coupled to the clarity of the constructs' mental representation that participants should rate. The number of categories to evaluate simulator sickness and digital eye strain ranges from four (SSQ [46]) to 100 (visual analog scale [86]) or even "an extremely strong pain that a person has ever experienced" (Borg CR10 [11]). Following Krosnick and Presser's argumentation we chose a 7-point scale, as this provides enough clarity for each individual category, while keeping a suitable differentiation of the single points on the scale [58]. This conclusion is supported by Menold and Bogner, who recommended to use rating scales of 5- or 7-points [75]. Symptom severity is typically measured with unipolar rating scales [3, 11, 46]. As the opposite of experiencing a symptom is not conceivable (but would rather refer to not experiencing the symptom) a bipolar rating scale is not suitable. While the number of categories of the scales we build upon varies between 4 and 100, they are all unipolar scales (e.g., [40, 86, 95]). Furthermore, verbally labelling categories increases test-retest reliability [75]. Typically, symptom severity or pain scales are labelled from "none" to "severe", but differ in labelling the single categories in between [3, 24, 45]. Other labels are "no discomfort"/"very bad discomfort" [40] or "nothing"/"very much" [95]. We chose to stay consistent with typical symptom severity scales but decided to verbalize all seven categories, as this was found to facilitate participants' mental representation of the measured construct. The scale we used was labelled as: "0

Table 2: The ergonomic symptoms included in our online study – based on Knight and Baber's CRS [54].

Ergonomic Symptoms	
Emotion	I was worried about how I look when I wear this device. I felt tense or on edge because I was wearing the device.
Attachment	I could feel the device on my body. I could feel the device moving.
Harm	The device was causing me some harm. The device was painful to wear.
Perceived Change	Wearing the device made me feel physically different. I felt strange wearing the device. I felt bulky wearing the device.
Movement	The device affected the way I move. The device inhibited or restricted my movement.
Anxiety	I did not feel safe wearing the device.
	I was not able to move as usual. I felt the device was too heavy. The attachment of the device was too tight.
Additional Statements	The attachment of the device was too loose. I felt that I did not have the device properly attached.
	The device generated additional heat leading to excess sweating.

(nothing at all)", "1 (very slight)", "2 (slight)", "3 (moderate)", "4 (moderately severe)", "5 (severe)", and "6 (very severe)". Inspired by the Borg CR10 scale and to increase comprehensibility, we added a description to the two poles. The maximal pain or discomfort that we expect to occur is a pain that makes participants stop or abort the experience. Consequently, the maximum was labelled "very severe (I don't want to use the device under these conditions)". We deliberately chose not to label it as "I want to abort or stop using the experience", as the questionnaire is employed after the experience. The minimum was labelled as "nothing at all (I don't experience this [symptom] at all".

4.2 Study Design and Procedure

The study was conducted as a between-subject design with the factor *administration practice*. That is, one group answered the survey post-exposure only (G1), and one group answered it before and after the exposure (G2). Participants were asked to fill out the post-exposure survey after the next time they would use their headset for more than 30 consecutive minutes. Given that we aimed for participants reflecting on their natural behaviour, we did not make restrictions about the application or VR experience they used. Using the same experience for all participants would have prevented us from achieving that goal and would likely have limited ecological validity of our experiment due to distorted occurrence of symptoms compared to natural usage. In addition and similar to Law et al. [63], we were interested in the factor structure of the symptom categories independently of the application. Participants of G2 had to fill out a pre-exposure questionnaire in addition to the post-exposure questionnaire. The pre-exposure questionnaire consisted of the two symptom categories simulator sickness and digital eye strain. The procedure was the same as for the post-exposure questionnaire described in the following:

After providing informed consent, participants were introduced to the rating scales and the three categories of symptoms. The order of the categories as well as the items were randomized within each participant. After answering the items of all three categories, participants were asked to rate the relative relevance that each category had to them with regard to discomfort in VR HMDs. To define this, we created all possible pairs ($n = 3$) and asked participants to choose the most relevant symptoms group to discomfort. This procedure was inspired by Hart and Staveland's method to define participants' subjective perception of workload [32]. Lastly, participants were asked to describe their recent VR experience, their usage behavior, and whether they had any visual impairments or vision problems. Participants of G1 took on average 12 minutes to complete the post-exposure questionnaire and received a compensation of £1.5 (£7.5 per hour). Participants of G2 took on average 4 minutes for the pre-exposure and 14 for the post-exposure questionnaire. They received a total compensation of £2.13 (£7.1 per hour). All questionnaires are available as supplementary material.

4.3 Participants

Participants were recruited via Prolific[4]. During registration, participants were asked about how frequently, for how long, and since when they used their VR headset. They were also asked about their usual experience with discomfort and whether they considered themselves susceptible to discomfort in VR. The following additional personal data was provided by the survey platform: age, sex, country of birth, country of current residence, employment status, first language, nationality, and student status. Only participants who agreed to receive an invitation to the second (G1) or second and third (G2) part of the study, were sent one. The registration phase took approx. two minutes (G1= 2.05, G2= 2.19). A total of 642 (G1=226, G2=333) participants were recruited and 559 completed all parts of the study. Of these 207 (G1=44, G2=163) were excluded due to several reasons described below. The final set of participants (352) consisted of 49% female and 51% male participants, with 182 for G1 (47% female, 53% male) and 170 for G2 (51% female, 49% male). The mean age of participants in G1 was 29 ($SD = 8, range : 18 - 50$) and 30 in G2 ($SD = 9, range : 18 - 54$).

4.3.1 Data Quality and Exclusion of Participants. Although it was observed that data quality in online studies is comparably reliable to conventional methods [14], it is important to apply a sensible inspection of the quality of the obtained data. Therefore, four reliability measures were defined to be able to exclude low-quality responses quickly. First, a number of attention checks were integrated into the pre- and post-exposure questionnaires (2 for pre-exposure, 3 for post-exposure), which were consistent with the survey platform's guidelines on fair attention checks[5]. Participants that missed one or more attention checks were excluded from analysis (G1=10, G2=42), but also received compensation. Secondly, three symptoms (blurred vision, eye strain, and difficulty focusing) were included twice (once

[4]https://prolific.co/ accessed: September 10th, 2020
[5]https://researcher-help.prolific.co/hc/en-gb/articles/360009223553-Using-attention-checks-as-a-measure-of-data-quality, accessed September 10th, 2020

in the simulator sickness section and once in the DES section) in the questionnaire to evaluate internal consistency of the answers. Responses were excluded that differed by two or more points of measurement in these symptoms (G1=5, G2=15). Thirdly, responses were excluded when participants indicated that they had used their headset for less than 30 minutes, or they claimed that they had used it for more than 30 minutes, but it was clear that they did not, given the time between registration and at which the questionnaires were completed (G1=29, G2=67). Finally, participants of G2 that completed the pre- and post-exposure questionnaire in the wrong order were excluded (39).

4.4 Results

4.4.1 *Integrating SSQ, DES, and ERG Into One Factor Model.*

Quality and reliability assessment. We first conducted a confirmatory factor analysis (CFA) to compare the simulator sickness data of the post-exposure survey of both groups with the SSQ 3-factor model, in order to obtain a first estimate of the quality and reliability of our data. The symptoms were specified to load on their designated factor as described by Kennedy et al. [46] and the factor loading of the first indicator was fixed to one. As our data were ordinal and the criterion of multivariate normality was not met (Mardia test [72]: $Skewness = 4407.5, p < .01, Kurtosis = 57.7, p < .01$), we used diagonally weighted least squares as estimation method [64, 76]. Results of the CFA suggest that the model is a good fit for the data ($\chi^2_{(96,N=352)} = 172.66, p < .001$ criteria $p > .05$ [39], $CFI = .994$ criteria $> .95$ [16], $TLI = .993$ criteria $> .95$ [39], $RMSEA = .048$ criteria $< .05$ [70], $p\text{-}RMSEA = .616$). The only criteria that indicated a rejection of the model was the χ^2 measure, which does assume multivariate normality that was not given in our case and is therefore plausible to result in an erroneous rejection of the model [74]. As an additional validity measure we employed a hierarchical factor analysis that allows to more deeply investigate the loadings of single items to respective general and group factors (similar to procedure in SSQ). The resulting structure showed great similarity to the original factor model of the SSQ, the details of which are available as supplementary material. In conclusion the results suggest that the overall structure of our simulator sickness data meets the proposed factor model of the SSQ.

Exploratory factor analysis of all three symptom categories. We conducted an exploratory factor analysis (EFA), following the guidelines summarized by Samuels [83], to build a factor model of general discomfort including the three categories simulator sickness, digital eye strain, and ergonomics. The analysis was conducted based on polychoric correlation matrices that were shown to provide more accurate reproductions of the measurement model for ordinal data than Pearson correlations [38]. Barlett's test that indicates whether correlations among the variables are present was significant ($\chi^2_{(2016,N=352)} = 25604.46, p < .001$) [5]. First, we conducted an EFA with principal axis factoring (PA) and varimax rotation to generate orthogonal factors. However, the average within factor correlation of the result (0.49) was only marginally higher than the average between factor correlation (0.42), which was not considered a satisfactory solution. Therefore, we repeated the analysis with oblimin rotation. The following steps were repeated and the analysis rerun until the solution stabilised.

First, parallel analysis was conducted to define the number of factors that should be extracted – 10 in our case. Therefore, we ran an EFA with PA, oblimin rotation, and 10 factors. Second, items with communality (h2) smaller than 0.3 were removed to ensure that all items shared some common variance with other items. The communality of an item is a measure of the proportional variance in that item that is explained by the extracted factors (in contrast to uniqueness (u2)). Then, items that had no factor loading > 0.3 were removed, followed by items that had cross-loadings with a maximal factor loading < 0.4.

A stable solution was yielded after 19 repetitions during which 24 items were removed (including duplicate items), resulting in a final set of 39 items. During analysis, factors were removed that had less than three items with loading > 0.4, except one factor that continuously had only two items (SS7, ERG117) with high factor loadings (> 0.7), which we kept for the final solution. As final result of the EFA we obtained 6 factors that are comprised by the items presented in Table 3. The results for each iteration and the details of the statistical validation tests that are reported in the following are available as supplementary material.

A solution is considered valid when there are higher average correlations between the items of the extracted factors (within factor correlation) than the average correlations between the factors (between factor correlation). To calculate the within and between factor correlation, we ran a principal component analysis (PCA) with one component for each of the factors, i.e., input for the PCA were the items that loaded on the respective factor. The within factor correlations were calculated for each factor using Pearson correlations. To calculate the between factor correlations the regression scores of the PCA were extracted for each factor. Based on these, a correlation matrix indicating the relation between factors was extracted. The average within factor correlation was with 0.53 higher than the between factor correlation of 0.4, which indicates an acceptable solution. The extracted communalities were all > 0.3 with $34 > 0.5$ and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA) was very good with Overall MSA = 0.87 (criteria >0.7 for good fit [17]). As a double-check, we ran a Cronbach's alpha reliability analysis on each factor, which also resulted in satisfactory values, as a value of > 0.7 is considered reliable for internal consistency ($\alpha_{f1} = 0.93, \alpha_{f2} = 0.9, \alpha_{f3} = 0.86, \alpha_{f4} = 0.83, \alpha_{f5} = 0.83, \alpha_{f6} = 0.77$) [77]. The six factors explained 61% of the variance in the data, which is also acceptable. In addition, a CFA confirmed that the model fitted the data well ($\chi^2(764, N = 352) = 1473.722, p < .001, CFI = .990, TLI = .990, RMSEA = .051, p\text{-}RMSEA = .269$).

4.4.2 *Comparison of Symptoms of Simulator Sickness, DES, and ERG.*

The six factors that were extracted in the previous step were applied to the post-exposure values of both groups and the pre-exposure values of G2 (see Figure 3, left for mean values and right for distribution of scores). Tests of normality (Shapiro-Wilk test and Komogorov-Smirnov test) were applied to all variables before analysing within- or between-group differences. More detailed information on test results are provided as supplementary material. As all factors were assessed with the same scale, a comparison indicating symptom severity in each factor was made with Wilcoxon

Table 3: The final factor structure of EFA with 39 respective items loading on 6 factors. For each factor the explained proportional amount of variance is listed. The items and factors are assigned with one of three codes that defines which symptom category they belong to. SS: simulator sickness symptoms, DES: digital eye strain symptoms, ERG: ergonomic symptoms.

Item	Factor	F1	F2	F3	F4	F5	F6	h2	u2	com
Irritation of eyes	Digital Eye Strain	**0.87**	-0.02	0.01	-0.01	0.02	-0.04	0.74	0.26	1.0
Soreness of eyes	Digital Eye Strain	**0.85**	-0.05	-0.01	0.03	0.01	0.06	0.72	0.28	1.0
Eye strain	Digital Eye Strain	**0.78**	0.04	-0.07	0.10	-0.05	0.10	0.69	0.31	1.1
Sensation of hot eyes	Digital Eye Strain	**0.78**	0.00	0.05	-0.09	0.16	0.00	0.68	0.32	1.1
Dry eyes	Digital Eye Strain	**0.75**	-0.06	-0.01	0.13	-0.05	-0.02	0.58	0.42	1.1
Burning eyes	Digital Eye Strain	**0.74**	0.20	-0.01	0.02	-0.09	0.09	0.73	0.27	1.2
Eye ache	Digital Eye Strain	**0.74**	0.05	0.08	-0.09	0.12	-0.03	0.65	0.35	1.1
Eye redness	Digital Eye Strain	**0.72**	0.15	0.06	0.02	-0.12	0.07	0.69	0.31	1.2
Tearing eyes	Digital Eye Strain	**0.67**	0.07	0.04	0.06	-0.12	0.06	0.55	0.45	1.1
Discomfort on eyes	Digital Eye Strain	**0.62**	-0.1	0.11	0.09	0.16	-0.14	0.52	0.48	1.4
Blurred vision	Digital Eye Strain	**0.43**	0.30	0.08	0.07	0.04	0.04	0.56	0.44	2.0
Stomach awareness	Simulator Sickness	-0.16	**0.69**	0.05	0.05	0.09	0.13	0.51	0.49	1.2
Nausea	Simulator Sickness	-0.04	**0.68**	-0.03	0.05	0.16	0.09	0.57	0.43	1.2
Dizzyness (open eyes)	Simulator Sickness	0.16	**0.66**	0.01	0.08	0.04	-0.02	0.65	0.35	1.2
Vertigo	Simulator Sickness	0.16	**0.61**	0.05	-0.01	0.08	0.04	0.58	0.42	1.2
Dizzyness (closed eyes)	Simulator Sickness	0.26	**0.54**	0.06	0.03	0.10	-0.01	0.62	0.38	1.6
Difficulty concentrating	Simulator Sickness	0.25	**0.52**	0.08	0.06	0.11	-0.13	0.61	0.39	1.8
Burping	Simulator Sickness	0.16	**0.49**	0.18	0.24	-0.33	0.14	0.65	0.35	3.1
Fullness of head	Simulator Sickness	0.20	**0.44**	0.10	0.04	0.11	-0.04	0.46	0.54	1.7
Difficulty focusing	Simulator Sickness	0.29	**0.43**	0.05	0.14	0.09	-0.11	0.57	0.43	2.4
General discomfort	Simulator Sickness	0.29	**0.32**	0.05	0.10	0.23	-0.05	0.51	0.49	3.1
Discomfort on shoulders	Neck/Shoulder Pain	0.03	-0.06	**0.93**	-0.05	-0.02	-0.01	0.80	0.20	1.0
Shoulder pain	Neck/Shoulder Pain	0.03	-0.01	**0.88**	0.00	-0.05	0.05	0.79	0.21	1.0
Discomfort on neck	Neck/Shoulder Pain	-0.03	-0.07	**0.82**	0.12	0.03	-0.01	0.70	0.30	1.1
Neck pain	Neck/Shoulder Pain	-0.05	0.14	**0.78**	0.00	0.08	0.01	0.72	0.28	1.1
Device not properly attached	Erg. (Attachment)	0.03	-0.02	0.02	**0.75**	0.02	0.01	0.61	0.39	1.0
Attachment too loose	Erg. (Attachment)	-0.09	0.10	-0.01	**0.74**	-0.03	0.04	0.54	0.46	1.1
Device painful to wear	Erg. (Attachment)	0.00	-0.06	0.19	**0.64**	0.16	-0.03	0.61	0.39	1.3
Could feel device moving	Erg. (Attachment)	0.01	0.10	-0.07	**0.62**	0.04	0.06	0.46	0.54	1.1
Device too heavy	Erg. (Attachment)	0.11	-0.14	0.04	**0.58**	0.22	0.04	0.55	0.45	1.5
Device causing some harm	Erg. (Attachment)	0.04	0.25	0.15	**0.52**	-0.04	-0.03	0.54	0.46	1.7
Attachment too tight	Erg. (Attachment)	0.26	0.04	0.00	**0.41**	0.04	0.05	0.39	0.61	1.8
Not able to move as usual	Erg. (Perceived Change)	0.02	0.06	0.03	0.04	**0.74**	-0.01	0.63	0.37	1.0
Device affected movement	Erg. (Perceived Change)	0.04	0.06	-0.01	0.07	**0.72**	0.10	0.66	0.34	1.1
Device restricted movement	Erg. (Perceived Change)	0.01	0.14	0.08	0.08	**0.68**	0.10	0.68	0.32	1.2
Feeling of device on body	Erg. (Perceived Change)	-0.06	0.04	-0.01	0.18	**0.46**	0.09	0.33	0.67	1.5
Feeling physically different	Erg. (Perceived Change)	0.14	0.21	0.07	0.09	**0.41**	-0.02	0.43	0.57	2.0
Device generated heat	Erg. (Sweating)	0.13	-0.14	-0.05	0.14	0.12	**0.79**	0.74	0.26	1.2
Sweating	Erg. (Sweating)	-0.05	0.14	0.13	-0.12	-0.01	**0.77**	0.63	0.37	1.2
Sums of Squares		8.32	4.82	3.6	3.73	2.92	1.5			
% of variance explained		20	12	9	9	7	4			

Signed Rank tests with Bonferroni correction. Tests were significant between all factors, except between F2/F3, F2/F4. and F3/4. Mean values of F5 (Perceived Change) ($M = 1.76, SD = 1.07$) and F6 (Sweating) ($M = 1.51, SD = 1.39$) were highest, followed by F1 (DES) ($M = 1.17, SD = 0.99$). Factors F2 (Simulator Sickness) ($M = 0.88, SD = 0.82$), F3 (Neck/Shoulder Pain) ($M = 0.88, SD = 1$), and F4 (Attachment) ($M = 0.92, SD = 0.85$) resulted in almost equal mean values. For factors F1, F5, and F6, approx. 30% of participants had at least moderate discomfort values, while this percentage was at 15% for factors F2, F3, and F4. The two symptoms that occurred least often, were *salivation increasing* with 78% not experiencing it at all and *burping* with 72% of participants not experiencing it. The symptom with the highest single symptom score was *eye strain* with 50% of participants having stated that it occurred at least slightly (15% moderately, 9% moderately severe or higher). This was followed by three statements for which at least 30% of the participants

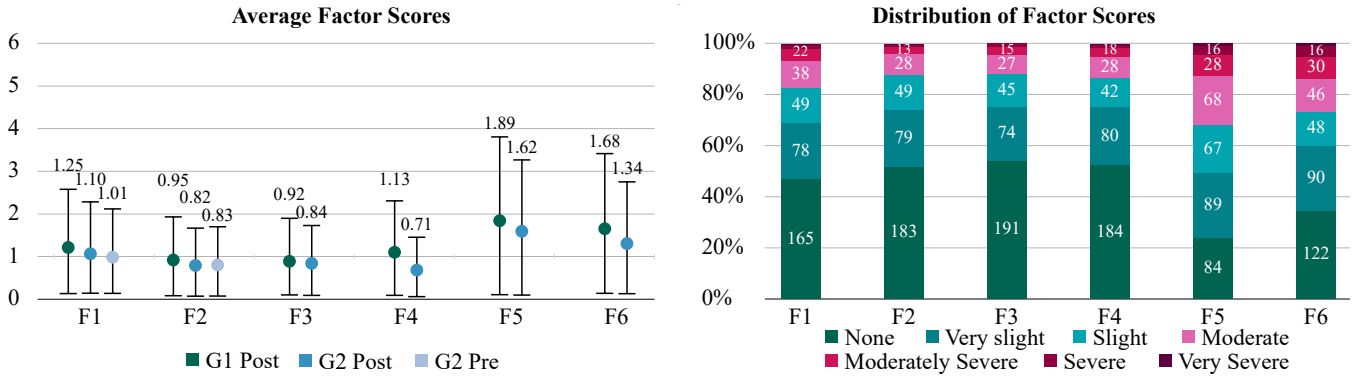

Figure 3: Left: mean factor scores, averaged over all participants. Right: distribution of factor scores for each factor, averaged over all symptoms that contribute to its respective factor.

rated them as at least moderate. These are "I could feel the device on my body" (34%), "the device was causing me some harm" (30%), and "the attachment of the device was too loose" (30%). Of the 9 symptoms that were rated on the adapted Borg CR10 scale, measuring perceived discomfort on different parts of the face, head, and neck, the highest rated symptoms were eyes and temples (for both 28% of participants rated them as moderate at least) (see Figure 4 for mean values).

4.4.3 Influences of Sex, Frequency of Usage, and Motion Conflict on Symptom Factors. Literature suggests influences of sex [26], frequency of usage [34], and visuo-vestibular conflict [45, 62] on simulator sickness. To reveal possible similar effects on our factor values, we conducted an univariate analysis of variances (ANOVA) on each factor (6 dependent variables F1-F6) with four between-subject variables *group* (G1/G2), *sex* (female/male), *frequency of usage* (0: less than once a week, 1: once a week, 2: once/twice a week, 3: daily, 4: several times a day), and *visuo-vestibular conflict* (yes/no). Although significance tests for normality were negative, Q-Q-plots and boxplots of the dependent variables indicated that data does not largely differ from normality (plots are provided as supplementary material). In addition, it was shown that ANOVA analysis is robust for non-parametric data of our sample size [8], which is why we considered it valid to proceed with the analysis. Homogenity of variances was given (tested with Levene's tes, $p > .05$). We found a significant effect of *group* on F2 ($F(1) = 4.92, p < .05$), F4 ($F(1) = 15.66, p < .01$), and F5 ($F(1) = 8.62, p < .01$) with G1 having slightly higher scores than G2 ($M(F2_{G1}) = 0.95, M(F2_{G2}) = 0.82$, $M(F4_{G1}) = 1.13, M(F4_{G2}) = 0.71$, $M(F4_{G1}) = 1.89, M(F4_{G2}) = 1.62$). Furthermore, we found a significant effect of *sex* on F2 ($F(1) = 7.33, p < .01$, univariate test: $F(1) = 8.38, p < .01$), i.e., on average women reported higher simulator sickness values ($M = 1.03, SD = 0.89$) than men ($M = 0.74, SD = 0.72$). We also found a significant effect of *frequency* on F5($F(4) = 2.42, p < .05$, univariate test: $F(4) = 2.74, p < .05$). Pairwise comparisons showed ($p < .05$) a significant difference between users that used their headset several times a day ($M = 1.36, SD = 0.94$) and users that used their headset less than once a week ($M = 2.08, SD = 1.18$). We did not find a significant effect for *visuo-vestibular conflict*, indicating that the type of experience did not have an influence on the symptoms results. However, we observed a trend for F2 ($F(1) = 3.67, p = .056$), indicating that values were higher ($M = 0.96, SD = 0.90$) when a conflict was present than when not ($M = 0.77, SD = 0.69$).

4.4.4 Administration Practice. As in the previous section, an effect of group was found on three factors (F2, F4, and F5), we calculated SSQ scores to reveal potential differences in administration practice of the questionnaire and to be able to compare our results to previous work on SSQ data. To calculate SSQ scores, we applied the following scale transformation: 0->1, 1->1, 2->1, 3->2, 4->2, 5->3, 6->3. First, potential differences between post-exposure SSQ scores (nausea, oculomotor, disorientaiton, and total score) between both groups were tested. A Mann-Whitney-U test did not reveal any significant differences between G1 group and G2 for none of the scores ($U_N = 14805, U_O = 14114, U_D = 13641, U_{TS} = 14049, p > .05$). Second, there were no statistically significant differences between pre- and post-exposure values of G2 (tested with Wilcoxon matched-pair signed rank test $W_N = 4261, W_O = 4851, W_D = 4020, W_{TS} = 5813, p > .05$).

4.4.5 Descriptive Results. The average usage time was 53 minutes ($SD = 34, range : 30 - 300$). This matched the usual usage time that participants indicated: most of them usually used their headset for 30-60 minutes (G1: 50%, G2: 49%), followed by 1-2 hours (G1: 29%, G2: 30%), less than 30 minutes (G1: 12%, G2: 14%), 2-3 hours (G1: 7%, G2: 4&), and more than 3 hours (G1/G2: 2%). Most of the participants had actively been using their headset for 1-6 months (G1: 29%, G2: 19%) or 6-12 months (G1: 29%, G2: 32%), followed for 1-2 years (G1: 22%, G2: 30%), more than 2 years (G1/G2: 16%), and only 3% (G1) and 2% (G2) had only used it for 1 month or less. The majority of participants usually used their headset for once (G1: 19%, G2: 23%) or twice a week (G1 38%:, G2: 39%). Some used it for once (G1: 18%, G2: 14%) or several times a day (G1: 14%, G2: 8%). 11% (G1) and 16% (G2) of participants used it less often than once a week. 40% (G1) and 25% (G2) played a video, 86% (G1) 85% (G2) played a game, 15% (G1) and 9% (G2) engaged in a social and 9% (G1/G2) in a creative experience and 14% (G1) and 9% (G2) browsed to a store or web page. 60% of G1 and 59% of G2 engaged in an experience where virtual motion was present. 42% of G1 and 45% of

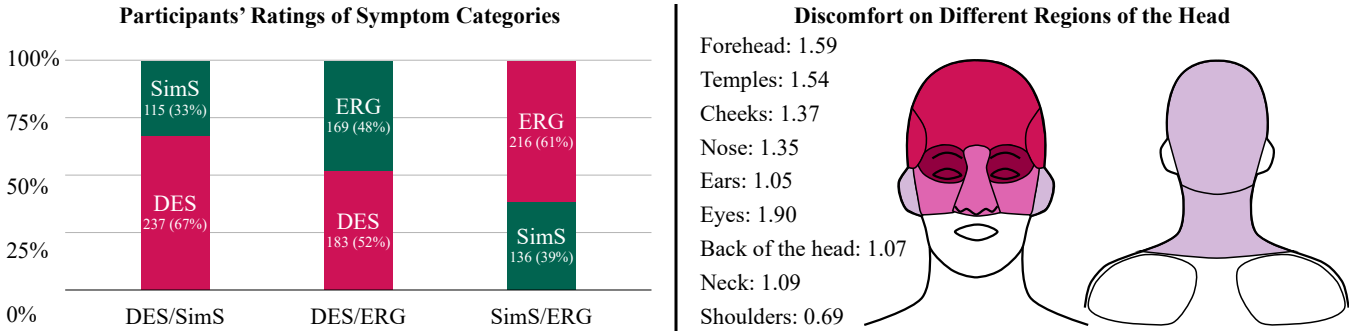

Figure 4: Left: Participants' perceived relevance of three symptom categories ergonomics (ERG), digital eye strain (DES), and simulator sickness (SimS) to general discomfort in VR HMDs. Right: Discomfort that participants perceived on different regions of their head.

G2 reported to have a vision problem and 40% of G1, 42% G2 used prescription glasses or contact lenses on a regular basis.

28 participants (8%) reported to use comfort features, such as additional cushions, counter weights, or face padding to increase comfort while wearing the headset. During registration, participants were asked to rate their experience with and susceptibility to discomfort. 117 participants (33%) reported to usually experience at least moderate discomfort when using their HMD (G1: 69/38%, G2: 48/28%). This number was slightly higher when rating their susceptibility, where 133 participants (38%) indicated that they would consider themselves at least moderately susceptible to experiencing discomfort in VR HMDs (G1: 67/37%, G2: 66/39%). After the exposure participants were asked to rate how relevant they considered each category of symptoms with regards to general discomfort by directly comparing each two of the categories, i.e., in total they rated three comparisons. In the comparisons of DES with simulator sickness and ERG with simulator sickness symptoms, in both cases simulator sickness was named by fewer participants (see Figure 4 left). In the comparison of DES and ERG, both were rated as roughly equally important.

5 DISCUSSION

5.1 On the Factor Model

In contrast to previous works that aimed to revise the factor structure of SSQ [4, 12], we aimed for building a more comprehensive model of discomfort, addressing DES and ERG in addition to simulator sickness. We identified several discrepancies when comparing our model with previous works: The symptom *headache* that was included in other factor structures [46, 49, 89] was removed during our refinement process. In contrast, *sweating* was kept although removed for revised factor structure versions of the SSQ [85]. Interestingly, the symptom *general discomfort* was attributed to the simulator sickness factor (F2). This may suggest that simulator sickness includes symptoms that address a more general feeling of discomfort in contrast to very specific symptoms, such as sweating or tearing eyes, which contribute to the other two symptom categories. Similar to Ames et al. [3] and in contrast to other previous works [86], *double vision* was not a relevant indicator of ocular symptoms and was removed from the factor model. This is particularly surprising, as the majority of participants did not use a HMD with adjustable inter pupillary distance. One reason might be that users were used to wearing and engaging with the device, as they used it very frequently and might not notice specific symptoms after long usage times. Our single factors match in large parts with previous work on simulator sickness and digital eye strain [46, 86]. However, we are the first to propose an orthogonal factor structure of discomfort in VR HMDs that includes three categories of symptoms, demonstrating their specific relations to each other.

5.2 On the Comparison of Symptoms of SSQ, DES, and ERG

The extracted factor model of discomfort in VR HMDs is split into six factors, four of which refer to ergonomic symptoms, explaining 29% of the variance in the data, followed by one DES factor accounting for 20%, as well as a simulator sickness factor explaining 12%. As important ergonomic factors, we identified shoulder and neck pain (F3), symptoms that relate to the attachment of the device (F4), symptoms that relate to perceived change (F5), and sweating (F6). Taken together, these results not only suggest that discomfort in VR HMDs is comprised by (at least) three main components (ERG, DES, and simulator sickness) but also that, among these three simulator sickness – although widely measured – seems to be the least important.

The distribution of scores in Figure 3 on the respective factors shows that about one third of the participants experienced at least moderate symptoms for F1-F4, and moderately severe to very severe values for F5 and F6. The symptom category with the highest severity scores was ERG, particularly F5 (perceived change) and F6 (sweating). Symptoms of F5 refer to the perceived change when wearing the device, including statements like "the device inhibited or restricted my movements" or "I was not able to move as usual". F6 consists of only two statements that cover sweating. In conclusion, the most often occurring symptoms are that participants feel different when using the device and start sweating when using it. The last one is possibly linked to the VR experience a person engages with. However, we found high sweat values across all participants that engaged in a variety of different applications, which speaks for wearing the device making people sweating. The

symptom category with the second highest prevalence was digital eye strain (F1), with the highest single symptom score of eye strain. Lastly, the category with lowest prevalence was simulator sickness (F2), with highest single item values of general discomfort. The two rarest symptoms "increased salivation" and "burping" were reported by less than 22% of the participants. Yet, burping was still included in the factor model. The distribution of factor scores and the values of each category are in agreement with participants' rating of perceived relevance of symptoms to general discomfort. When comparing DES symptoms with ERG symptoms, participants' ratings are almost balanced. However, when comparing both with symptoms of simulator sickness, participants rate simulator sickness symptoms as less relevant in both cases (Figure 4 left). This is a strong indicator that if researchers are aiming to employ some form of measure for general discomfort, using the SSQ is not enough and does not include other partially more relevant symptoms. One suggestion can be to either clarify what type of discomfort is expected and select the corresponding questionnaire, or employ all three factors (ERG, DES, and simulator sickness) to get a more holistic understanding of potential occurrences of discomfort.

5.3 On the Severity of Symptoms

We found a mean total SSQ score of 25 in the reviewed papers ($range : 0 - 235.62$). These findings are in agreement with Stanney et al., who report a mean total score of 29 in eight VE studies [88]. However, mean total SSQ scores of our user study were almost twice as high with an average post-exposure score of 53. In contrast to Stanney et al., who argue that experienced users of flight simulators experience less severe symptoms, our results indicate the contrary, as we had a sample of participants that frequently used their headset [88]. Additionally, we found that users who used their headset several times a day had lower simulator sickness values than users that used their headset more rarely, which supports the hypothesis of an habituation effect.

Peculiarly, the pre-exposure value of our user study was at 50. These findings indicate that our sample had already high symptoms before starting the experiment. This might be explained by conducting a user survey with an online study portal, where users presumably spend a long time in front of screens per day and therefore might have increased symptoms due to prolonged screen time. However, we did not measure screen time and can therefore only hypothesise about this cause. In addition, participants were frequent VR users who may experience symptoms, but less severe than they could be measured with a 7-point scale. We rather presume that the VR exposure of 30 minutes was too short to induce a significant change in symptoms that could be assessed with the rating scales, when starting with already high symptoms. Another explanation for the high pre-exposure values might be that we measured symptoms on a 7-point scale, in contrast to SSQ that has a 4-point scale. To analyse SSQ values, we transformed data to the SSQ scale. It was shown that 7-point scales provide a clearer understanding of the measured construct for participants than shorter scales [58]. Given that we provided more categories on a fine-grained scale, participants possibly reported higher values than they would have with a 4-point scale. Therefore, the data-transformation might have caused higher values than when measured directly with a 4-point scale. This might also serve as explanation for the high pre-exposure values. Lastly, Kennedy et al. classified symptoms that occur in simulator studies of > 15 as a concern [44]. Although Stanney et al. argued that cybersickness results in other symptom distributions and higher values than simulator sickness [88], it can be questioned if a questionnaire that indicates a problem at 15 on a scale from $0 - 235.62$ is fine-grained enough to measure small differences in symptomatology. In summary, these findings indicate that a 7-point scale allows for assessing more detailed differences in single scale categories than using a 4-point scale (SSQ). However, although we used a more specific scale than SSQ, both the original and the extended scales are still too insensitive to measure fine differences in perceived symptoms.

5.4 On the Administration Practice of Discomfort Scales

Our review shows that there is no clear consensus in the community about whether the SSQ should be used as an absolute (only post-exposure) or relative (pre- and post-exposure) measure. 60% of the papers presented the questionnaire post-hoc, the other 40% before and after an experiment. These findings are in line with prior work where arguments both in favor [3] and against [94] both approaches can be found. Comparing both ways of administration of the questionnaire, we did not find statistically significant differences between post-exposure SSQ values of both groups in our user study. These findings are in agreement with Ames et al. [3], who also found no significant differences of post-exposure values when employing it post-exposure only versus employing pre- and post-exposure. However, these findings stand in contrast to Young et al., who found that repeatedly answering the SSQ results in higher values than employing it only post-exposure [94]. Although we cannot propose a clear solution based on our results on administration practice, we suggest to use relative values (pre- and post-exposure). Pre-exposure values could help to calibrate symptomatology data that is assessed with severity scales, establishing a none-zero baseline. An open question remains, whether high pre-exposure values are due to participants' social desirability bias [30].

5.5 On the Generalizability of the Sample

We found a large gender bias in the reviewed papers with only 39% female participants. This represents another limitation of simulator sickness research that should be addressed in the future. In our user study we found that women rated simulator sickness symptoms higher than men ($M_{women} = 1.03, M_{men} = 0.74$). While it has been argued that this is rather due to susceptibility than an effect of sex [20, 26], we did not observe a similar effect for any other factor (which were all measured with the same scale). We therefore conclude, like other works before [2, 7, 27], that it is rather likely that women experience higher values of simulator sickness than men.

5.6 Limitations and Future Work

In the first part of this work we presented a systematic literature review, investigating the current use of SSQ. Among other, we collected the authors' rationales of employing the questionnaire. When

no specific reasons for employment were given, we concluded that authors have used the SSQ as a standard measure of sickness or discomfort symptoms in VR. This procedure might have introduced a bias to the results. Furthermore, although we included three scientific databases in our review, we might have missed works that were published at other venues, and therefore cannot claim to present a holistic review of all SSQ research.

In the second part of this work, we have identified and shown the composition of three main components that contribute to discomfort in VR. While we were able to further subdivide one of them (ERG) into four sub factors, DES and simulator sickness are each comprised by 11 and 10 items, which we could not further refine in our work. However, in SSQ as well as in different DES questionnaires, a further separation was proposed [46, 86]. Another limitation of our work is that we only investigated three symptom categories of discomfort. In addition, these addressed physiological symptoms. VR can cause psychological problems, such as claustrophobia [73], that are not addressed by our current model and have to be further investigated and potentially integrated into an even more comprehensive discomfort model in future. Further, the user study was conducted with frequent users of VR HMDs, with most of the participants using their headset at least once a week. This may have led to an overall low increase in symptom values. Therefore, it will be interesting to repeat the study with novices or users that are not as accustomed to VR devices, as this may yield higher symptom values. This may also shed some light on whether experienced users are less prone to experiencing symptoms or whether the rating scale is indeed too insensitive to measure small differences in symptoms.

6 CONCLUSION

Using a systematic literature review, in this work we uncovered the widespread practice of using the SSQ as a measure of general discomfort in VR research – although the questionnaire was not originally developed for this purpose and although it only covers one of several prevalent and serious, but largely overlooked, symptom categories. We therefore proposed and studied an extended factor model of discomfort in VR HMDs that includes digital eye strain and ergonomics in addition to simulator sickness. Using a large-scale online study, we found that symptoms caused by headsets' ergonomics were indeed most prevalent, and together with digital eye strain, affecting users the most. Similar to Kennedy et al. [46], who at the time developed the SSQ in response to a new technology (simulators), our findings call not only for a change in the common use of SSQ today but also for developing a more comprehensive questionnaire for measuring general discomfort. We do not claim to already present such a questionnaire, but hope that our work will trigger a discussion on both topics to improve future research on discomfort on VR head-mounted displays.

ACKNOWLEDGMENTS

This work has been supported by the DFG projects "Gaze-Assisted Scalable Interaction in Pervasive Classrooms" (RU 1605/5-1) and "Empirical Assessment of Presence and Immersion in Augmented and Virtual Realities" (RU 1605/4-1) and by the Landesgraduiertenförderung (LGFG) Scholarship for PhD students.

REFERENCES

[1] Fedaa Abd-Alhamid, Michael Kent, John Calautit, and Yupeng Wu. 2020. Evaluating the impact of viewing location on view perception using a virtual environment. *Building and Environment* 180 (2020), 106932. https://doi.org/10.1016/j.buildenv.2020.106932

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

[3] Shelly L. Ames, James S. Wolffsohn, and Neville A. McBrien. 2005. The Development of a Symptom Questionnaire for Assessing Virtual Reality Viewing Using a Head-Mounted Display. *Optometry and Vision Science* 82, 3 (2005), 571–572. https://doi.org/10.1097/01.OPX.0000156307.95086.6

[4] Stacy A. Balk, Mary Anne Bertola, and Vaughan W. Inman. 2013. Simulator sickness questionnaire: twenty years later. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. University of Iowa, Bolton Landing, New York, USA, 257–263. https://doi.org/10.17077/drivingassessment.1498

[5] Maurice S. Bartlett. 1954. A note on the multiplying factors for various χ 2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)* 16, 2 (1954), 296–298. https://doi.org/10.1111/j.2517-6161.1954.tb00174.x

[6] Nabil Benzina, Margot Morgiève, Marine Euvrard, João Flores Alves Dos Santos, Antoine Pelissolo, and Luc Mallet. 2020. Personalised 360o video exposure therapy for the treatment of obsessive-compulsive disorder: A single case study. *French Journal of Psychiatry* 1 (2020), 31 – 38. https://doi.org/10.1016/j.fjpsy.2020.02.004

[7] Maximino Bessa, Miguel Melo, David Narciso, Luís Barbosa, and José Vasconcelos-Raposo. 2016. Does 3D 360 Video Enhance User's VR Experience? An Evaluation Study. In *Proceedings of the XVII International Conference on Human Computer Interaction* (Salamanca, Spain) *(Interacción '16)*. ACM, New York, NY, USA, Article 16, 4 pages. https://doi.org/10.1145/2998626.2998669

[8] María J. Blanca, Rafael Alarcón, Jaume Arnau, Roser Bono, and Rebecca Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema* 29, 4 (2017), 552–557.

[9] Clayton Blehm, Seema Vishnu, Ashbala Khattak, Shrabanee Mitra, and Richard W. Yee. 2005. Computer Vision Syndrome: A Review. *Survey of Ophthalmology* 50, 3 (2005), 253 – 262. https://doi.org/10.1016/j.survophthal.2005.02.008

[10] Benjamin Bolte, Gerd Bruder, Frank Steinicke, Klaus Hinrichs, and Markus Lappe. 2010. Augmentation Techniques for Efficient Exploration in Head-Mounted Display Environments. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology* (Hong Kong) *(VRST '10)*. ACM, New York, NY, USA, 11–18. https://doi.org/10.1145/1889863.1889865

[11] Gunnar Borg. 1998. *Borg's perceived exertion and pain scales.* Human Kinetics, Champaign, IL, US. Pages: viii, 104.

[12] Stéphane Bouchard, Geneviève Robillard, and Patrice Renaud. 2007. Revising the factor structure of the Simulator Sickness Questionnaire. *Annual review of cybertherapy and telemedicine* 5, Summer (2007), 128–137.

[13] Susan Bruck and Paul A. Watters. 2011. The factor structure of cybersickness. *Displays* 32, 4 (2011), 153 – 158. https://doi.org/10.1016/j.displa.2011.07.002 Visual Image Safety.

[14] Michael D. Buhrmester, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon's Mechanical Turk. *Perspectives on Psychological Science* 6 (2011), 3 – 5. https://doi.org/10.1177/1745691610393980

[15] Jorge Cancela, Matteo Pastorino, Alexandros T. Tzallas, Markos G. Tsipouras, Giorgios Rigas, Maria T. Arredondo, and Dimitrios I. Fotiadis. 2014. Wearability assessment of a wearable system for Parkinson's disease remote monitoring based on a body area network of sensors. *Sensors (Basel, Switzerland)* 14, 9 (September 2014), 17235–17255. https://doi.org/10.3390/s140917235 Publisher: MDPI.

[16] Marianne Carlson and Stanley A. Mulaik. 1993. Trait Ratings from Descriptions of Behavior As Mediated by Components of Meaning. *Multivariate Behavioral Research* 28, 1 (1993), 111–159. https://doi.org/10.1207/s15327906mbr2801_7 PMID: 26824998.

[17] Barbara A. Cerny and Henry F. Kaiser. 1977. A Study Of A Measure Of Sampling Adequacy For Factor-Analytic Correlation Matrices. *Multivariate Behavioral Research* 12, 1 (1977), 43–47. https://doi.org/10.1207/s15327906mbr1201_3 PMID: 26804143.

[18] Eunhee Chang, Hyun Taek Kim, and Byounghyun Yoo. 2020. Virtual Reality Sickness: A Review of Causes and Measurements. *International Journal of Human–Computer Interaction* 0, 0 (2020), 1–25. https://doi.org/10.1080/10447318.2020.1778351

[19] Wei Chen, Jian-Gang Chao, Xue-Wen Chen, Jin-Kun Wang, and Cheng Tan. 2015. Quantitative orientation preference and susceptibility to space motion sickness simulated in a virtual reality environment. *Brain Research Bulletin* 113 (2015), 17 – 26. https://doi.org/10.1016/j.brainresbull.2015.01.007

[20] Wei Chen, Jian-Gang Chao, Xue-Wen Chen, Jin-Kun Wang, and Cheng Tan. 2015. Quantitative orientation preference and susceptibility to space motion sickness

simulated in a virtual reality environment. *Brain Research Bulletin* 113 (2015), 17 – 26. https://doi.org/10.1016/j.brainresbull.2015.01.007

[21] Lorenzo Tancredi D'Angelo, Jan Parlow, Wolfgang Spiessl, Stefan Hoch, and Tim C. Lüth. 2010. A system for unobtrusive in-car vital parameter acquisition and processing. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, New York, NY, USA, 1–7. https://doi.org/10.4108/ICST.PERVASIVEHEALTH2010.8877

[22] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. 2014. A Systematic Review of Cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment* (Newcastle, NSW, Australia) *(IE2014)*. ACM, New York, NY, USA, 1–9. https://doi.org/10.1145/2677758.2677780

[23] Thomas W. Day and Nigel W. John. 2019. Training Powered Wheelchair Manoeuvres in Mixed Reality. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. IEEE, New York, NY, USA, 1–7. https://doi.org/10.1109/VS-Games.2019.8864515

[24] María del Mar Seguií, Julio Cabrero-García, Ana Crespo, José Verdú, and Elena Ronda. 2015. A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace. *Journal of Clinical Epidemiology* 68, 6 (2015), 662 – 673. https://doi.org/10.1016/j.jclinepi.2015.01.015

[25] Ajoy S. Fernandes and Steven K. Feiner. 2016. Combating VR sickness through subtle dynamic field-of-view modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, New York, NY, USA, 201–210. https://doi.org/10.1109/3DUI.2016.7460053

[26] Moira B. Flanagan, James G. May, and Thomas G. Dobie. 2005. Sex Differences in Tolerance to Visually-Induced Motion Sickness. *Aviation, Space, and Environmental Medicine* 76, 7 (2005), 642–646. https://www.ingentaconnect.com/content/asma/asem/2005/00000076/00000007/art00004

[27] Sebastian Freitag, Benjamin Weyers, and Torsten W. Kuhlen. 2016. Examining Rotation Gain in CAVE-like Virtual Environments. In *IEEE Transactions on Visualization and Computer Graphics*. IEEE, New York, NY, USA, 1462–1471. https://doi.org/10.1109/TVCG.2016.2518298

[28] Vivian Genaro Motti and Kelly Caine. 2014. Understanding the Wearability of Head-Mounted Devices from a Human-Centered Perspective. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers* (Seattle, Washington) *(ISWC '14)*. ACM, New York, NY, USA, 83–86. https://doi.org/10.1145/2634317.2634340

[29] Guilherme Gonçalves, Miguel Melo, and Maximino Bessa. 2018. Virtual Reality Games: A Study About the Level of Interaction vs. Narrative and the Gender in Presence and Cybersickness. In *2018 International Conference on Graphics and Interaction (ICGI)*. IEEE, New York, NY, USA, 1–8. https://doi.org/10.1109/ITCGI.2018.8602686

[30] Pamela Grimm. 2010. Social Desirability Bias. In *Wiley International Encyclopedia of Marketing*. John Wiley & Sons, Inc., New York, NY, USA. https://doi.org/10.1002/9781444316568.wiem02057

[31] Jens Grubert, Yuta Itoh, Kenneth Moser, and John Edward Swan. 2018. A Survey of Calibration Methods for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 24, 9 (2018), 2649–2662. https://doi.org/10.1109/TVCG.2017.2754257

[32] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. Elsevier, North-Holland, 139 – 183. https://doi.org/10.1016/S0166-4115(08)62386-9

[33] Lawrence J. Hettinger, Kevin S. Berbaum, Robert S. Kennedy, William P. Dunlap, and Margaret D. Nolan. 1990. Vection and Simulator Sickness. *Military Psychology* 2, 3 (1990), 171–181. https://doi.org/10.1207/s15327876mp0203_4 PMID: 11537522.

[34] Katherine J. Hill and Peter Alan Howarth. 2000. Habituation to the side effects of immersion in a virtual environment. *Displays* 21, 1 (2000), 25 – 30. https://doi.org/10.1016/S0141-9382(00)00029-9

[35] Teresa Hirzle, Maurice Cordts, Enrico Rukzio, and Andreas Bulling. 2020. A Survey of Digital Eye Strainin Gaze-Based Interactive Systems. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. ACM, New York, NY, USA, Article 9, 12 pages. https://doi.org/10.1145/3379155.3391313

[36] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 625, 12 pages. https://doi.org/10.1145/3290605.3300855

[37] Jimmy Ho, Reinhard Pointner, Huai-Chun Shih, Yu-Chih Lin, Hsuan-Yu Chen, Wei-Luan Tseng, and Mike Y. Chen. 2015. EyeProtector: Encouraging a Healthy Viewing Distance when Using Smartphones. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. ACM, New York, NY, USA, 77–85. https://doi.org/10.1145/2785830.2785836

[38] Francisco Pablo Holgado–Tello, Salvador Chacón–Moscoso, Isabel Barbero–García, and Enrique Vila–Abad. 2008. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity* 44, 1 (September 2008), 153. https://doi.org/10.1007/s11135-008-9190-y

[39] Daire Hooper, Joseph Coughlan, and Michael R. Mullen. 2008. Structural Equation Modelling: Guidelines for Determining Model Fit. *ElectronicJournal of Business Research Methods* 6, 1 (2008), 53–60.

[40] Peter Alan Howarth and Howell Owen Istance. 1985. The association between visual discomfort and the use of visual display units. *Behaviour & Information Technology* 4, 2 (1985), 131–149. https://doi.org/10.1080/01449298508901794

[41] Ian T. Joliffe and Bjt Morgan. 1992. Principal component analysis and exploratory factor analysis. *Statistical Methods in Medical Research* 1, 1 (1992), 69–95. https://doi.org/10.1177/096228029200100105 PMID: 1341653.

[42] David B. Kaber, Yulan Liang, Yu Zhang, Meghan L. Rogers, and Shruti Gangakhedkar. 2012. Driver performance effects of simultaneous visual and cognitive distraction and adaptation behavior. *Transportation Research Part F: Traffic Psychology and Behaviour* 15, 5 (2012), 491 – 501. https://doi.org/10.1016/j.trf.2012.05.004

[43] Robert S. Kellogg, Robert S. Kennedy, and Ashton Graybiel. 1964. Motion Sickness Symptomatology of Labyrinthine Defective and Normal Subjects During Zero Gravity Maneuvers. Technical Documentation Report NO. AMRL-TDR-64-47. , 1 pages.

[44] Robert S. Kennedy, Julie M. Drexler, Daniel E. Compton, Kay M. Stanney, D. Susan Lanham, and Deborah L. Harm. 2003. Configural scoring of simulator sickness, cybersickness, and space adaptation syndrome: Similarities and differences. In *Virtual and adaptive environments: Applications, implications, and human performance issues*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 247–278. https://doi.org/10.1201/9781410608888.ch12

[45] Robert S. Kennedy, Julie M. Drexler, and Robert C. Kennedy. 2010. Research in visually induced motion sickness. *Applied Ergonomics* 41, 4 (2010), 494 – 503. https://doi.org/10.1016/j.apergo.2009.11.006

[46] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (1993), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

[47] Behrang Keshavarz and Heiko Hecht. 2011. Validating an Efficient Method to Quantify Motion Sickness. *Human Factors* 53, 4 (2011), 415–426. https://doi.org/10.1177/0018720811403736 PMID: 21901938.

[48] Do Hoe Kim, Donald E. Parker, and Min Young Park. 2004. A New Procedure for Measuring Simulator Sickness–the RSSQ. , 14 pages.

[49] Hyun K. Kim, Jaehyun Park, Yeongcheol Choi, and Mungyeong Choe. 2018. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics* 69 (2018), 66 – 73. https://doi.org/10.1016/j.apergo.2017.12.016

[50] Joohwan Kim, David Kane, and Martin S. Banks. 2014. The rate of change of vergence–accommodation conflict affects visual discomfort. *Vision Research* 105 (2014), 159 – 165. https://doi.org/10.1016/j.visres.2014.10.021

[51] James F. Knight and Chris Baber. 2004. Neck Muscle Activity and Perceived Pain and Discomfort Due to Variations of Head Load and Posture. *Aviation, Space, and Environmental Medicine* 75, 2 (2004), 123–131. https://www.ingentaconnect.com/content/asma/asem/2004/00000075/00000002/art00005

[52] James F. Knight and Chris Baber. 2005. A Tool to Assess the Comfort of Wearable Computers. *Human Factors* 47, 1 (2005), 77–91. https://doi.org/10.1518/0018720053653875 PMID: 15960088.

[53] James F. Knight and Chris Baber. 2007. Effect of Head-Mounted Displays on Posture. *Human Factors* 49, 5 (2007), 797–807. https://doi.org/10.1518/001872007X230172 PMID: 17915598.

[54] James F. Knight, Chris Baber, Anthony Schwirtz, and Huw W. Bristow. 2002. The comfort assessment of wearable computers. In *Proceedings. Sixth International Symposium on Wearable Computers,*. IEEE, New York, NY, USA, 65–72. https://doi.org/10.1109/ISWC.2002.1167220

[55] James F. Knight, Daniel Deen-Williams, Theodoros N. Arvanitis, Chris Baber, Sofoklis Sotiriou, Stamatina Anastopoulou, and Michael Gargalakos. 2006. Assessing the Wearability of Wearable Computers. In *2006 10th IEEE International Symposium on Wearable Computers*. IEEE, New York, NY, USA, 75–82. https://doi.org/10.1109/ISWC.2006.286347

[56] Frank C. Koslucher, Eric J. Haaland, and Thomas A. Stoffregen. 2014. Body load and the postural precursors of motion sickness. *Gait & Posture* 39, 1 (2014), 606 – 610. https://doi.org/10.1016/j.gaitpost.2013.09.016

[57] Andrey Krekhov, Sebastian Cmentowski, Katharina Emmerich, and Jens Krüger. 2019. Beyond Human: Animals as an Escape from Stereotype Avatars in Virtual Reality Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 439–451. https://doi.org/10.1145/3311350.3347172

[58] Jon A. Krosnick and Stanley Presser. 2010. Question and Questionnaire Design. *Handbook of survey research* 94305, 886 (2010), 1432–1033.

[59] Kenichiro Kutsuna, Yasuyuki Matsuura, Kazuhiro Fujikake, Masaru Miyao, and Hiroki Takada. 2013. Comparison of form in potential functions while maintaining upright posture during exposure to stereoscopic video clips. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine*

and Biology Society (EMBC). IEEE, New York, NY, USA, 7200–7203. https://doi.org/10.1109/EMBC.2013.6611219

[60] Luis Antonio L. F. da Costa and Luciana Porcher Nedel. 2017. An Immersive Visualization Study on Molecules Manipulation. In *2017 19th Symposium on Virtual and Augmented Reality (SVR)*. IEEE, New York, NY, USA, 169–177. https://doi.org/10.1109/SVR.2017.30

[61] Chengyuan Lai, Ryan P. McMahan, and James Hall. 2015. March-and-Reach: A realistic ladder climbing technique. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, New York, NY, USA, 15–18. https://doi.org/10.1109/3DUI.2015.7131719

[62] Joseph J. LaViola. 2000. A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bull.* 32, 1 (Jan. 2000), 47–56. https://doi.org/10.1145/333329.333344

[63] Effie L.-C. Law, Florian Brühlmann, and Elisa D. Mekler. 2018. Systematic Review and Validation of the Game Experience Questionnaire (GEQ) - Implications for Citation and Reporting Practice. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. ACM, New York, NY, USA, 257–270. https://doi.org/10.1145/3242671.3242683

[64] Cheng-Hsien Li. 2016. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods* 48, 3 (Sept. 2016), 936–949. https://doi.org/10.3758/s13428-015-0619-7

[65] Alessandro Liberati, Douglas G. Altman, Jennifer Tetzlaff, Cynthia Mulrow, Peter C. Gøtzsche, John P. A. Ioannidis, Mike Clarke, P. J. Devereaux, Jos Kleijnen, and David Moher. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology* 62, 10 (2009), e1 – e34. https://doi.org/10.1016/j.jclinepi.2009.06.006

[66] Yun Ling, Willem-Paul Brinkman, Harold T. Nefs, Chao Qu, and Ingrid Heynderickx. 2012. Effects of Stereoscopic Viewing on Presence, Anxiety, and Cybersickness in a Virtual Reality Environment for Public Speaking. *Presence* 21, 3 (2012), 254–267. https://doi.org/10.1162/PRES_a_00111

[67] Christophe Lopez, Bigna Lenggenhager, and Olaf Blanke. 2010. How vestibular stimulation interacts with illusory hand ownership. *Consciousness and Cognition* 19, 1 (2010), 33 – 47. https://doi.org/10.1016/j.concog.2009.12.003

[68] Astrid J.A. Lubeck, Jelte E. Bos, and John F. Stins. 2015. Motion in images is essential to cause motion sickness symptoms, but not to increase postural sway. *Displays* 38 (2015), 55 – 61. https://doi.org/10.1016/j.displa.2015.03.001

[69] Jean-Luc Lugrin, Marc Cavazza, David Pizzi, Thurid Vogt, and Elisabeth André. 2010. Exploring the Usability of Immersive Interactive Storytelling. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology* (Hong Kong) *(VRST '10)*. ACM, New York, NY, USA, 103–110. https://doi.org/10.1145/1889863.1889887

[70] Robert C. MacCallum, Michael W. Browne, and Hazuki M. Sugawara. 1996. Power analysis and determination of sample size for covariance structure modeling. *Psychological methods* 1, 2 (1996), 130.

[71] Divine Maloney. 2019. [DC] Embodied Virtual Avatars and Potential Negative Effects on Implicit Racial Bias. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 1373–1374. https://doi.org/10.1109/VR.2019.8798008

[72] Kantilal Vardichand Mardia. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57, 3 (1970), 519–530. https://doi.org/10.1093/biomet/57.3.519

[73] Hanna McCabe-Bennett, Richard Lachman, Todd A. Girard, and Martin M. Antony. 2020. A Virtual Reality Study of the Relationships Between Hoarding, Clutter, and Claustrophobia. *Cyberpsychology, Behavior, and Social Networking* 23, 2 (2020), 83–89.

[74] Cameron N. McIntosh. 2007. Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007). *Personality and Individual Differences* 42, 5 (2007), 859 – 867. https://doi.org/10.1016/j.paid.2006.09.020

[75] Natalja Menold and Kathrin Bogner. 2016. *Design of Rating Scales in Questionnaires (Version 2.0)*. GESIS - Leibniz-Institut für Sozialwissenschaften, Mannheim, Germany. 13 pages. https://doi.org/10.15465/gesis-sg_en_015

[76] Diana Mîndrilă. 2010. Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal of Digital Society* 1, 1 (2010), 60–66. https://doi.org/10.20533/ijds.2040.2570.2010.0010

[77] Jum C. Nunnally. 1994. *Psychometric theory 3E*. Tata McGraw-hill education, New York, NY, USA.

[78] Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A. Cooper, and Gordon Wetzstein. 2017. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences* 114, 9 (2017), 2183–2188. https://doi.org/10.1073/pnas.1617251114

[79] Monika Pölönen, Toni Järvenpää, and Jukka Häkkinen. 2010. Comparison of Near-to-Eye Displays: Subjective Experience and Comfort. *Journal of Display Technology* 6, 1 (2010), 27–35. https://doi.org/10.1109/JDT.2009.2029123

[80] Lisa Rebenitsch and Charles Owen. 2016. Review on cybersickness in applications and visual displays. *Virtual Reality* 20, 2 (June 2016), 101–125. https://doi.org/10.1007/s10055-016-0285-9

[81] Mark Rosenfield. 2011. Computer vision syndrome: a review of ocular causes and potential treatments. *Ophthalmic and Physiological Optics* 31, 5 (2011), 502–515. https://doi.org/10.1111/j.1475-1313.2011.00834.x

[82] Mark Rosenfield. 2016. Computer vision syndrome (aka digital eye strain). *Optometry* 17, 1 (2016), 1–10.

[83] Peter Samuels. 2017. Technical Report on ResearchGate: Advice on exploratory factor analysis. , 7 pages. https://www.researchgate.net/publication/319165677_Advice_on_Exploratory_Factor_Analysis

[84] Dimitrios Saredakis, Ancret Szpak, Brandon Birckhead, Hannah A. D. Keage, Albert Rizzo, and Tobias Loetscher. 2020. Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Frontiers in human neuroscience* 14 (March 2020), 96–96. https://doi.org/10.3389/fnhum.2020.00096 Publisher: Frontiers Media S.A.

[85] Volkan Sevinc and Mehmet Ilker Berkman. 2020. Psychometric evaluation of Simulator Sickness Questionnaire and its variants as a measure of cybersickness in consumer virtual environments. *Applied Ergonomics* 82 (2020), 102958. https://doi.org/10.1016/j.apergo.2019.102958

[86] James E. Sheedy, John Hayes, and Jon Engle. 2003. Is all Asthenopia the Same? *Optometry and Vision Science* 80, 11 (2003), 8. https://journals.lww.com/optvissci/Fulltext/2003/11000/Is_all_Asthenopia_the_Same_.8.aspx

[87] Alexis D. Souchet, Stéphanie Philippe, Dimitri Zobel, Floriane Ober, Aurélien Lévêque, and Laure Leroy. 2018. Eyestrain Impacts on Learning Job Interview with a Serious Game in Virtual Reality: A Randomized Double-blinded Study. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. ACM, New York, NY, USA, Article 15, 12 pages. https://doi.org/10.1145/3281505.3281509

[88] Kay M. Stanney, Robert S. Kennedy, and Julie M. Drexler. 1997. Cybersickness is Not Simulator Sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 41, 2 (1997), 1138–1142. https://doi.org/10.1177/107118139704100292

[89] William B. Stone III. 2017. *Psychometric Evaluation of the Simulator Sickness Questionnaire as a Measure of Cybersickness*. Ph.D. Dissertation. Iowa State University. https://doi.org/10.31274/etd-180810-5050

[90] Wibirama Sunu, Wijayanto Titis, Nugroho Hanung, Bahit Muhammad, and Winadi Mumtaz N. 2015. Quantifying visual attention and visually induced motion sickness during day-night driving and sleep deprivation. In *2015 International Conference on Data and Software Engineering (ICoDSE)*. IEEE, New York, NY, USA, 191–194. https://doi.org/10.1109/ICODSE.2015.7436996

[91] Ancrêt Szpak, Stefan Carlo Michalski, Dimitrios Saredakis, Celia S. Chen, and Tobias Loetscher. 2019. Beyond Feeling Sick: The Visual and Cognitive Aftereffects of Virtual Reality. *IEEE Access* 7 (2019), 130883–130892. https://doi.org/10.1109/ACCESS.2019.2940073

[92] Hiroki Takada and Matsuura Yasuyuki. 2013. Comparison of Form in Potential Functions While Maintaining Upright Postures during Exposure to Stereoscopic Video Clips. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, New York, NY, USA, 2140–2145. https://doi.org/10.1109/SMC.2013.366

[93] Caglar Yildirim. 2020. Don't make me sick: investigating the incidence of cybersickness in commercial virtual reality headsets. *Virtual Reality* 24, 2 (June 2020), 231–239. https://doi.org/10.1007/s10055-019-00401-0

[94] Sean D. Young, Bernard D. Adelstein, and Stephen R. Ellis. 2007. Demand Characteristics in Assessing Motion Sickness in a Virtual Environment: Or Does Taking a Motion Sickness Questionnaire Make You Sick? *IEEE Transactions on Visualization and Computer Graphics* 13, 3 (2007), 422–428. https://doi.org/10.1109/TVCG.2007.1041

[95] Fabrizio Zeri and Stefano Livi. 2015. Visual discomfort while watching stereoscopic three-dimensional movies at the cinema. *Ophthalmic and Physiological Optics* 35, 3 (2015), 271–282. https://doi.org/10.1111/opo.12194