Applied Ergonomics 82 (2020) 102958

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

# Psychometric evaluation of Simulator Sickness Questionnaire and its variants as a measure of cybersickness in consumer virtual environments

Volkan Sevinc[a], Berkman Mehmet Ilker[b,*]

[a] Bahcesehir University, Institute of Social Sciences, Game Design Graduate Programme, Turkey
[b] Bahcesehir University, Faculty of Communication, Dept. of Communication Design, Turkey

ARTICLE INFO

*Keywords:*
Cybersickness
Virtual reality
Visually induced motion sickness
Simulator sickness questionnaire
Scale
Psychometric evaluation

ABSTRACT

Cybersickness, i.e. visually induced motion sickness, remains as a negative effect that is detrimental to the user experience of VEs (virtual environments) developed for VR (virtual reality) consumers. As the VR technology evolves, it is rather triggered by application aspects rather than hardware limitations. For this reason, there is still a need for a measurement method to assess and compare VEs for cybersickness effects.

SSQ (Simulation Sickness Questionnaire) is used for measuring users' level of sickness symptoms and is highly appreciated in VR research. However, it is criticized for its psychometric qualities and applicability in VR, as a measure of cybersickness. Recently, two variants of SSQ were offered for measuring cybersickness, CSQ (Cybersickness Questionnaire) and VRSQ (Virtual Reality Sickness Questionnaire). There is also another variant with a different factor structure, which we call FSSQ, that is based on French translation of SSQ.

Our study compares SSQ and these variants for their psychometric qualities; construct validity, discriminant validity, internal reliability, test-retest reliability and sensitivity to distinguish application aspects of VEs that are related to cybersickness. Using a within-subjects experiment design, we evaluated 7 different VEs with 32 participants through 9 sessions, resulting with 288 responses to the 16-item SSQ. Results suggested that both VRSQ and CSQ were valid and reliable measures of cybersickness, as well as being sensitive to application aspects such as translational and rotational movements required by users for navigation in VEs. Compared to SSQ and FSSQ; the cybersickness questionnaires, CSQ and VRSQ, revealed better indicators of validity. On the other hand, we assume that the development of the two cybersickness scales had limitations in sample size to represent VR consumers and limitations in stimuli to represent the applications aspects of consumer VEs. We suggest further evaluation of cybersickness symptoms with larger samples and broader range of applications to identify the symptoms and the construct of a subjective measurement tool.

## 1. Introduction

Interacting with VR (virtual reality), especially with a HMD (head mounted display) equipped system, may trigger symptoms similar to motion sickness, referred to as visually induced motion sickness (Hettinger and Riccio, 1992) or cybersickness more suitably (McCauley & Sharkey,1992; LaViola, 2000). These negative effects are potentially detrimental for users' engagement with VEs (virtual environments) (Nichols and Patel, 2002). While HMDs became available as consumer products, manufacturers struggled with hardware limitations that might induce cybersickness, such as field of view, display resolution, refresh rate, flicker, temporal delays and input-output latency by enhancing hardware attributes such as binocular displays, interpupillary distance, position tracking sensors. As the commercial HMDs has overcome many of those hardware limitations today, there is a developing market of VE software for public use, especially for entertainment purposes and younger consumers are most likely to be interested in VR entertainment (Greenlight VR, 2015).

However; cybersickness, mainly based on application aspects rather than hardware limitations; remains as an issue that is potentially discomforting VR users, although there are VE design guidelines offered by researchers based on empirical evaluation (Porcino et al., 2017; Tanaka and Takagi, 2004; Jerald, 2015:209–213) in addition to some best practices suggested by developers according to professional experience (Google Design Guideline, n.d.; Oculus VR LLC, 2017; Epic Games, 2015). Therefore, there is still a need for methods to assess the cybersickness prior to the design and development of VEs.

Several methods are available to detect and measure cybersickness.

* Corresponding author.
*E-mail address:* berkman@comm.bau.edu.tr (M.I. Berkman).

https://doi.org/10.1016/j.apergo.2019.102958
Received 3 July 2018; Received in revised form 10 June 2019; Accepted 18 September 2019
Available online 26 September 2019
0003-6870/ 

Some measures are based on the aspects of VEs such as spatial velocity (So et al., 2001b); or based on physiological responses of subjects such as finger temperature (Nalivaiko et al., 2015), heart rate, respiratory rate, skin conductance, heterophoria (Rebenitsch and Owen, 2016); or based on cognitive or behavioral responses of subjects such as prolongation of reaction time (Nalivaiko et al., 2015) or postural instability (Chardonnet et al., 2017; Takada et al., 2007; Cobb, 1999). However, questionnaires, which are based on self-report responses of users, are the earliest method of assessment and "often the *de facto* rating systems of choice" (Rebenitsch and Owen, 2016).

Self-report measures of cybersickness in VEs have originated from motion sickness and simulator sickness studies. The Simulator Sickness Questionnaire (SSQ) (Lane and Kennedy, 1988) was obtained from the Motion Sickness Questionnaire (MSQ) (Kennedy and Graybiel., 1965). Re-analyzing MSQ data collected via simulator studies with participants employed in military, aviation and marine services; Lane and Kennedy (1988) determined 16 symptoms of motion sickness as psychometrically sound indicators of simulator sickness. SSQ symptoms indicate three constructs of simulator sickness: Nausea (N), Disorientation (D) and Oculomotor (O) effects, along with a second order more general factor as Total Severity (TS), which were largely adopted in later studies, although the factor structure of these symptoms causes some challenges in standardization (Kennedy et al., 2010).

Considering the issues caused by the complex factor structure of original SSQ, Bouchard et al. (2007) suggested a two-factor structure for SSQ, based on data compiled from non-military participants via the French translation of the 16 symptoms (F-SSQ). They criticized SSQ for being developed based on data collected through a population of highly trained professionals. Factors were revised as nausea and oculomotor based on data collected from 371 participants representing "a population of adults from the general public".

CSQ (Cybersickness Questionnaire) (Stone III, 2017) is an attempt to select the symptoms in SSQ that would appear to clearly indicate cybersickness, as some of the symptoms measured in SSQ might be triggered by cybersickness as well as other causes, such as sweating which might occur due to physical effort. Retained 9 symptoms indicated two factors: Dizziness and Difficulty in Focusing.

Kim et al. (2018) proposed the Virtual Reality Sickness Questionnaire (VRSQ), employing 9 symptoms of original SSQ to indicate Oculomotor and Disorientation constructs. Arguing that original SSQ was conceived from the data from simulator studies in early 90's, their study revised SSQ based on data collected via mobile HMD devices using four different stimuli in a within-subjects experiment design.

The original SSQ and two-factor variant F-SSQ were suggested as simulator sickness measures and the datasets used for the scale development process included several flight simulators besides VR systems. They both employ the same set of 16 symptoms while their factor structure and scoring methods are different. The two shorter forms, CSQ and VRSQ were proposed as cybersickness measures, developed based on the data acquired through participants who use HMD based VR systems. Each employ 9 symptoms selected amongst original 16 symptoms. 5 of these symptoms are common for both scales. CSQ uses a scoring method based on item weights suggested by its developer while VRSQ employs a simpler scoring method.

Our work aims to assess psychometric qualifications of original SSQ and its variants, F-SSQ, CSQ and VRSQ, for measuring cybersickness in publicly available HMD VR applications. Although SSQ is seen as a "gold standard" (Chan et al., 2010) for evaluation of visually induced motion sickness symptoms, we agree with the criticisms on its complex factor structure and development process that involves a sample of highly trained professionals and simulator experiments, which do not comply with the modern day HMD based VR and entertainment oriented users. Besides, scientists may benefit a shorter-form scale for assessment of cybersickness, as some researchers (eg. Alshaer et al., 2017; Aykent et al., 2014) tend to use a subset of SSQ items without conducting any psychometric evaluation.

The scales are compared for indicators of reliability, validity and sensitivity; through a dataset collected from a younger group of participants who are more likely to use VR for entertainment. Data is collected within experiments conducted with seven publicly available VE applications running on a consumer HMD.

Through comparison of psychometric qualities of SSQ and its variants based on our dataset, we aim to contribute to the subjective evaluation methods of cybersickness in publicly available HMD VR applications. Firstly, we want to compare each scale for their construct validity with a factor analytic approach in order to understand their applicability in for measuring cybersickness in publicly available HMD VR applications. Secondly, we aim to provide evidence for their internal consistency. Besides, we explore the changes on their results in repeated experiments while the users' experience increase, namely test-retest reliability. Finally, we examine their sensitivity to different application aspects, mainly navigation mechanisms that yield to different levels of scene complexity, which indicates scales' discriminant validity in evaluation of cybersickness in consumer VEs.

## 2. Related studies

### 2.1. Subjective evaluation methods of motion sickness symptoms

Pensacola Motion Sickness Questionnaire (MSQ) (Kellogg et al., 1965) is known to be the earliest survey based method of evaluating motion sickness in multiple dimensions, which evolved into Pensacola Diagnostic Index (Graybiel et al., 1968). Kennedy et al. (1994) suggested that simulator sickness is a different phenomenon from seasickness and MSQ is not a valid measure of simulator sickness as some of its symptoms, such as drowsiness which is a sopite symptom, are rarely observed in simulator studies. SSQ is developed based on analysis of MSQ data collected in simulator studies. Two variants of SSQ evaluated in our study, CSQ and VRSQ, were also based on an assumption that virtual reality induced symptoms of motion sickness are different from other types of simulators. Some studies rearranged the symptom sets for assessment of virtual reality systems (Ames et al., 2005) but SSQ remained a popular instrument in VR research. Another widely used alternative is Motion Sickness Assessment Questionnaire (MSAQ) (Gianaros et al., 2001). MSAQ consists of 20 symptoms representing four latent structures: gastro-intestinal, central, peripheral and sopite-related symptoms. Gastro-intestinal symptoms include being nauseated and queasy, and also refer to stomach. Central symptoms refer to dizziness and disorientation, being faint–like and lightheaded but also blurred vision is one of the symptoms. Peripheral symptoms refer to being sweaty, clammy/cold sweat and hot. Being annoyed, tired and having fatigue are sopite-related symptoms. The sopite-related symptoms are purposefully added to the scale based on the work of Graybiel and Knepton (1976) as the Pensacola Motion Sickness Questionnaire and Pensacola Diagnostic Index had fewer sopite-related items. Although MSAQ was developed to measure motion sickness in general without a special focus on VR and simulators, the development process is based on experiments that depend on visually-induced motion sickness via optokinetic drums. However, it had been employed in a simulator study (Brooks et al., 2010) and also in recent years it was employed in VR studies as an indicator of cybersickness (Speicher et al., 2018; Peterson et al., 2018; Wu et al., 2018; Nesbitt et al., 2017; Gavgani et al., 2017a). On the other hand, Gavgani et al. (2018) revealed that MSAQ symptoms are identical when same group of users were stimulated either via physical motion (Coriolis cross-coupling) or immersed in a VR roller-coaster, as well as electrodermal indicators associated with motion sickness, suggesting that cybersickness symptoms are not different from motion sickness symptoms in general.

There are also short evaluation methods to assess motion sickness as a unidimensional construct. Fast Motion Sickness Scale (FMS) (Keshavarz and Hecht, 2011b) emloys a 0 to 20 verbal rating scale of no-sickness/frank sickness which can be used quickly capture MS data during exposure. Results show that scores obtained using this measure highly correlate with SSQ dimensions and total severity scores (Keshavarz and Hecht, 2011b, 2014; Reinhard et al., 2017). Several studies also employed similar single-item assessment methods to measure simulator sickness level (e.g. McCauley et al., 1990; Helland et al., 2016) but unlike FMS, these methods were not psychometrically evaluated.

### 2.2. Other variants of SSQ

RSSQ (Revised simulation sickness questionnaire) is an expansion of SSQ and evaluates simulator sickness in 4 dimensions, using 24 items evaluated through an 11 point Likert scale (Kim et al., 2004). In addition to 16 SSQ symptoms, RSSQ employs 8 additional items that are for measuring drowsiness, visual flashbacks, awareness of breathing, confusion, vomiting, pallor, difficulty equilibrating, and muscle stiffness from strain. Besides Nausea, Oculomotor and Disorientation dimensions of SSQ, RSSQ incorporates a fourth dimension as “Strain/Confusion”. Transforming the 11 point RSSQ scores to 4 point SSQ scores, authors reported correlations > .7 between two scales' dimensions and total score.

Bruck and Watters (2011) offered a 4 factor model of cybersickness assessed by a combination of self-report and physiological measures. Employing SSQ symptoms, cardiac activity measurements, respiratory rate measurements and anxiety scores in their analysis, they identified the four factors as cybersickness, vision, arousal and fatigue.

Although those scales are also variants based on SSQ, they employ additional manifest variables or different scale points. F-SSQ, CSQ and VRSQ are completely derived from SSQ symptoms.

The following part explains the psychometric qualities of SSQ, F-SSQ, CSQ and VRSQ and provides information about studies that explores the aspects of VEs related to cybersickness.

### 2.3. Psychometric qualifications of SSQ and its evaluated variants

Symptoms of SSQ were derived from an initial set of 28 MSQ items used to evaluate motion sickness symptoms in flight simulators (Kennedy et al., 1993). The evaluated flight simulators were 10 model types such as helicopter simulators with 6DoF (six-degrees-of-freedom) moving base, fixed-base, fixed-wing and dome display simulators (Kennedy et al., 1989, 1992) used in military training.

16 retained symptoms were selected according their frequency and severity in 1119 simulator tests, in which pre-exposure and post-exposure data were collected. Symptoms that showed systematic changes from pre-exposure to post-exposure were chosen.

Although some symptoms such as vomiting are clearly important signs of motion and simulator sickness, they were not retained since they were rarely observed. Some of the excluded symptoms like boredom had high frequency but considered as misleading since they were not observed consistently with other symptoms.

A series of varimax rotated principal component analyses revealed a three factor structure as the most interpretable solution based on retained 16 symptoms: Nausea (N), Oculomotor (O) and Disorientation (D). However, some symptoms generated considerable loads on multiple factors. By reason of assigning items that had a varimax loading > 0.30 as an indicator of a factor, some symptoms indicated multiple factors, as se-UPDATEen in Table 1. Each cluster of symptoms reflected an impact of simulator exposure on a different system within the human body. Transforming these three factors into a hierarchical structure, a superior general construct of simulation sickness is proposed as index of Total Severity (TS). The scores for each construct and the superior Total Severity (TS) score is calculated as follows, using sums of total scores A, B and C in Table 1.

$$N = [A] \times 9.54 O = [B] \times 7.58 D = [C] \times 13.92 TS = [A] + [B] + [C] \times 3.74$$

Majority of the studies that were conducted using the SSQ only reported the TS score (Lin et al., 2002; Draper et al., 2001; Lampton et al., 2000). Several studies provided evidence that SSQ and its subdimensions are sensitive to different display types (Garris-Reif and Franz, 1995; Häkkinen et al., 2002), system configurations (Cobb et al., 1999; Nelson et al., 2000) or simulator types (Kennedy et al., 1989, 1992, 1994). Different types of platforms, like helicopter, fixed wing, driving simulators and HMD, BOOM VR and CAVE VR systems have different SSQ profiles. Drexler (2006) reported an O > N > D profile for flight simulators while VR systems and driving simulators presented an O > D > N profile of SSQ subdimensions. She suggested a proportional scoring approach for profile comparison of subdimensions. In her method, each scale score was divided by sum of all three subscores.

Bouchard et al. (2007) criticized the SSQ for its complicated factor structure and for its being developed on a dataset compiled from military professionals who were not representative of the adults from of the general public. They used French translation of 16-item SSQ, administered to 307 participants (71% female) who evaluated different HMDs with different tracking systems and a CAVE-like VR environment, utilizing several tasks such as exposure to feared stimuli, attention and exploration. Using principal axis factoring with varimax rotation, they suggested a two-factor solution that minimized cross-loading of items. Comparing their two-factor model with the original three-factor model of SSQ using AIC (Akaike) and BIC (Bayes) information criteria of model fit comparison, the two factor model provided better evidence for construct validity. Bouchard et al. (2007) did not provide a clear administration and scoring for F-SSQ. They reported their results as mean values for Nausea and Oculomotor factors and total score as sum of these means. The [D] and [E] values in Table 1 are mean values of corresponding items, and total score is calculated as [D] + [E].

Stone III (2017) hypothesized that original factor model would not fit SSQ responses in state-of-the-art consumer-oriented VR applications. Three commercially developed VEs for HMDs were evaluated by 202 participants (37% female and 63% male with no prior VR experience) employing a between-subjects experimental design, using an HMD connected to a personal computer. He suggested that two of the VEs, which were a car racing game and a roller-coaster simulation, had discordant visual motion and vestibular information since the users move in the VEs by vehicular locomotion but the users' point of view is controlled by HMD position. The third VE provides a room-scale experience that locomotion is acquired by stepping in the actual room and point of the view is controlled by HMD tracking. Thus, the experience imparts a concordant visual motion and vestibular information. The data analysis based on an item-response theory approach failed to support the original three-factor structure of SSQ and the alternative models based on 16 symptoms. The author suggested that “could have been a result of a combination of sparsity (overall low symptom incidence) and inclusion of SSQ items that may not be indicative of cybersickness.” To adjust sparsity by low symptom incidence, the moderate symptoms and severe symptoms were amalgamated, i.e. rescoring the severe (3) scored items to moderate (2), which resulted in a scoring system of 0 (none), 1 (slight) and 2 (moderate). 9 items were selected among the SSQ items amongst the other alternatives and a two-factor structure (Dizziness and Difficulty in focusing) was offered which presented the best model fit. Factor scores were calculated as the sum of each amalgamated item score multiplied by its weight (see Table 1) as:

$$[F] = \text{sum}(\text{itemScore x itemWeight})$$

$$[G] = \text{sum}(\text{itemScore x itemWeight})$$

Those weights were based on an exploratory factor analysis on data collected by the author. No total score for a superior factor was offered, but it is suggested that “Dizziness” factor describes the high levels of cybersickness, while “Difficulty in Focusing” low levels.

Kim et al. (2018) proposed another different subset of 9 SSQ

**Table 1**
Symptoms in dimensions of simulator sickness and cybersickness scales.

| | SSQ | | | F-SSQ | | CSQ | | VRSQ | |
|---|---|---|---|---|---|---|---|---|---|
| | Nausea | Oculomotor | Disorientation | Nausea | Oculomotor | Dizziness | Difficulty in focusing | Oculomotor | Disorientation |
| General discomfort | ✓ | ✓ | | ✓ | | | | ✓ | |
| Fatigue | | ✓ | | | ✓ | | | ✓ | |
| Headache | | ✓ | | | ✓ | .50 | | | ✓ |
| Eyestrain | | ✓ | | | ✓ | | .58 | ✓ | |
| Difficulty focusing | | ✓ | ✓ | | ✓ | | .89 | ✓ | |
| Increased salivation | ✓ | | | ✓ | | | | | |
| Sweating | ✓ | | | ✓ | | | | | |
| Nausea | ✓ | | ✓ | ✓ | | .84 | | | |
| Difficulty concentrating | ✓ | | ✓ | | ✓ | | | | |
| Fullness of head | | | ✓ | | ✓ | | .55 | | ✓ |
| Blurred vision | | ✓ | ✓ | | ✓ | | .81 | | ✓ |
| Dizzy (eyes open) | | | ✓ | ✓ | | .89 | | | |
| Dizzy (eyes closed) | | | ✓ | ✓ | | .99 | | | ✓ |
| Vertigo | | | ✓ | ✓ | | .54 | | | ✓ |
| Stomach awareness | ✓ | | | ✓ | | | | | |
| Burping | ✓ | | | | ✓ | | | | |
| | [A] | [B] | [C] | [D] | [E] | [F] | [G] | [H] | [I] |

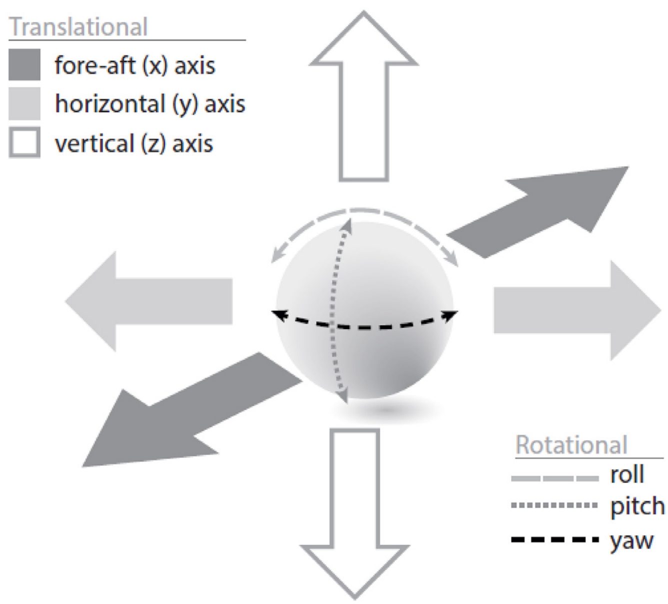


**Fig. 1.** Navigational axes in virtual environments. Adapted from Rebenitsch and Owen (2016).

symptoms driven by two latent constructs, “Oculomotor” and “Disorientation”, denoting SSQ dimension profiles are different for simulators and VR (Drexler, 2006; Kennedy et al., 2010). 5 symptoms were common with CSQ, but distributed to different factors. Among 12 male and 12 female participants, 6 had prior experience with HMDs and 3 owned a VR device. 4 different treatment conditions of button selection tasks were used as stimuli, employing 2 different selection methods and 2 different button sizes. 96 sessions were conducted with 24 participants under 4 different treatment conditions. Participants used a smartphone based HMD system for the tasks and responded to SSQ items after each treatment condition given in a Latin-square design, with 2 min breaks between each condition. An exploratory factor analysis using principal component analysis method with varimax rotation, 11 symptoms were retained with three components. However, the third component was indicated by only “Headache” symptom, which was removed. “Dizziness with eyes open” item was removed from “Disorientation” component due to high covariance value. A CFA (confirmatory factor analysis) verified the fit of the derived model. A simple averaging method was proposed to calculate scores for each component and a total index, in which [H] and [I] values were the sum of the corresponding items in Table 1. Analysis of variance (ANOVA) results revealed that there were significant differences in treatment conditions except for the disorientation score between two button size conditions, which was a static task that did not involve sudden movements. ANOVA results provided evidence for proposed VRSQ is sensitive to different application aspects, in terms of navigation and scene complexity. On the other hand, SSQ scores were found to be sensitive to application aspects for all dimensions.

Oculomotor = ([H] / 12) x 100Disorientation = ([I] / 15) x 100Total = (Oculomotor score + Disorientation score) / 2

### 2.4. Application Aspects Affecting Cybersickness

SSQ-T scores range from 6 to 160, with an average of upper twenties and “cybersickness appears to increase with more in-depth immersion and forced motion” (Rebenitsch and Owen, 2016). Besides hardware related issues such as field of view (Draper et al., 2001) or limited peripheral vision (Moss and Muth, 2011), application aspects in terms of navigation and scene complexity, rotation, independent visual backgrounds, tracking and position, scene content and realism are known to affect cybersickness (Rebenitsch and Owen, 2016)

Navigation mechanisms in VEs let user to move virtually in translational and rotational axes by generating moving scenes. Translational axes are fore-aft (x), horizontal (y) or vertical (z) and rotational axes are roll, pitch and yaw (Fig. 1). Current consumer VR HMD's provide rotational navigation by generating scenes that are opposite to direction of user's actual movement, where actual movement is detected via head tracking.

Exploring a VE with rotational navigation based on head movements is found to cause higher cybersickness compared to hand-controlled rotational navigation with stationary head position (Howarth and Finch, 1999).

On the other hand, translational navigation in VEs cannot be matched to user's actual movement accurately, since the users actually move in a limited space or do not move at all. Locomotion interfaces based on actual movements of the user, such as treadmills and movable tiles (Iwata, 2013) are not offered as consumer products yet. Some VEs offer “teleportation” mechanisms for navigation (Jerald, 2015:344). Yet the translational navigation mainly depends on input devices such as game controllers or some sort of automated solutions that are not controlled by the users. Users' bodies are stationary in the actual space while they are experiencing an illusory self-motion in virtual space, i.e.

vection. Vection was associated with cybersickness, in line with sensory mismatch theory (Rebenitsch and Owen, 2016), which is the most common theory of visually induced motion sickness.

So, Lo and Ho (2001a) explored the effect of navigation speed in the fore-aft direction and rotation in the yaw axis using a virtual urban scene. 10 male Chinese users reported higher sense of vection and higher SSQ-T scores when translational movement speed in fore-aft axis was increased from 3 m/s to 10 m/s and became steady after 10 m/s up to 59 m/s. SSQ and vection scores were not affected significantly by duration of immersion in VE. In an experiment with 7 male and 7 female participants, Bonato et al. (2008) reported significantly higher SSQ scores for alternating vection compared to a steady stimulus, within all dimensions of SSQ. So, Ho and Lo (2001b) offered a quantified metric of SV (spatial velocity) for perceived movement in VEs, in order to identify the relationship between navigation speed, scene content, and cybersickness. As a measure of scene complexity, SV was based on the luminance frequency in scene, which could be affected by the movements of users in translational and rotational axes. The cybersickness dose value model (So, 1999), which aimed to predict the severity of cybersickness that might be caused by immersion into a VE, was defined as the integral of SV over time multiplied by display, task, and individual scaling factors.

Dorado and Figueroa (2014) revealed that the cybersickness was lower for ramps compared to stairs due to smoother movement. Although increasing and decreasing the acceleration map real-world movement better, instant stop/start with velocity controlled by joystick tilt caused less motion sickness. Comparison of active (6DoF), active-passive (3DoF) and no-controls conditions resulted in lower cybersickness scores for 3DoF controls that matched the application purpose (Stanney et al., 2002). Serge and Mosh (2015) did not detect a significant difference in motion sickness between an observational task and a navigational task. However, participants who performed the observational task first scored significantly higher in navigational task, compared to participants who performed the navigational task first.

Watanabe and Ujike (2008) observed statistically significant lower SSQ-D scores for a movement in vertical axis alongside for-aft axis, compared to movement only in horizontal axis, but SSQ-T scores were statistically same for both conditions. Comparison of head/body position based controls and joystick controls yielded better results for head/body position tracking based control (Chen et al., 2013). Rebenitsch and Owen (2016) suggested that the axis of translational movement was not a contributing factor in navigation cybersickness. Instead, the effects were likely to lie with scene content and navigation control. Cybersickness was less likely with smooth navigation and moderate control.

When the movement in rotational axes were explored in terms of cybersickness, SSQ-N scores increased for all pitch, yaw, and roll axis rotations (So and Lo, 1999) but no axis yielded significantly higher scores than the other. Bonata et al. (2009) revealed that combined rotational movement yielded higher cybersickness. Keshavarz and Hecht (2011a, 2011b) had similar findings comparing single axis rotational movement to multiple axis rotations of users on a roller-coaster VE. With a rotating tunnel stimulus, Dennison and D'Zmura (2017) determined that SSQ scores did not differ significantly when participants were seated or standing, but cybersickness increased accordingly for both positions as rotation speed increased. Palmisano et al. (2017) used a rotational stimulus that was inversely-compensated to the same direction of users' head movements, which produced higher SSQ scores compared to compensated and uncompensated rotations.

## 3. Methods

### 3.1. Participants

Participants were volunteers from a group of students who had enrolled to courses offered by one of the researchers. A 15-min face-to-face briefing was given to candidates in classroom, describing the content of the VE's used in the study and possible side-effects of VR use. They were asked to consider their vulnerability to media such as horror movies and their prior experiences of motion sickness in transportation before agreeing to volunteer for the study. None of the candidates had prior VR experience. For this reason, volunteers were allowed to resign from the study after the first VR session. Of 41 initial participant group, 9 declared that they would not like to continue due to discomfort, 6 left due to horror content and 3 related due to cybersickness symptoms. The remaining 32 participants (17 females), 5 with corrected vision were aged from 19 to 25 (M = 21; SD = 1.91), who could be considered as possible young consumers of VR systems interested in cyber-entertainment.

### 3.2. Apparatus and stimuli

The apparatus was Oculus Rift Consumer Version 1 (CV1) HMD attached to a Windows 10 64-bit personal computer with an i7 processor and R9 90 series graphic card used for running HMD. CV1 has a field of view of 110°. Resolution is 1080x1200 per eye with image refreshing rate of 90 Hz. Logitech F310 gamepad was used as a joystick controller when necessary.

Seven publicly available VE applications were employed as stimuli which were horror-movie like experiences with suspended horror elements and jump scares. Some examples of suspended horror elements were audio effects such as squeaks or deep screams, blinking lights or representations of supernatural creatures. Jump scares were sudden changes in visual scene with appearance of unnatural objects, persons or creatures. Stimuli had a narrative theme in relation with supernatural events or violence. All experiences were in first person view, using head tracking for rotational movement. Mean experience durations are given in Table 2 along with the major navigation properties of applications.

Demonic Guest VR (Mgsstudio, 2016) was used as stimuli A1 and A2 in first and last sessions. The experience eventuates in a bedroom on a large bed but users do not see any representations of their body parts as an avatar. On the right side of the room, there is a closed circuit television screen that participant can view other rooms and there was a TV

**Table 2**
VE applications and experience durations.

| Code | Application Name | Translational Movement | Locomotion | Mean Exp. Time (Sec.) | SD |
|---|---|---|---|---|---|
| A1 | Demonic Guest | none | none | 364 | 41 |
| B | Abe VR | none | none | 361 | 35 |
| C1 | Final Rest | Fore axis | Vehicular - Auto | 213 | 20 |
| D | The Night The Carsons Disappeared | none | none | 1190 | 447 |
| E | Affected | 3DoF | Joystick controlled walking. | 740 | 335 |
| F | The Visitor | none | none | 582 | 202 |
| G | Insane Decay of Mind | 3DoF | Joystick controlled walking. | 527 | 230 |
| C2 | Final Rest B | Fore axis | Vehicular - Auto | 224 | 30 |
| A2 | Demonic Guest | none | none | 370 | 26 |

and window on the left. During the experience, users are induced to turn their head to look at the occurring jump scare event via audio and lighting cues. Users watch a demon on closed circuit monitor, approaching and finally reaching their room. There is no translational movement, but all interaction is based on movements on rotational axes, mainly yaw actions.

ABE VR (Hammerhead VR, 2016) was used as stimulus B on second session. Experience takes place in a medical operating room like environment. Users see the representation of their upper bodies and legs, arms and feet strapped to an operating table. A humanistic robot comes into the room and starts to tell a story about itself while it is walking beside the operating table and playing with surgical tools. Finally, it chops out the body parts of the user avatar. There is no translational movement but users are forced to rotate their heads to follow the robot, mainly with yaw actions.

Final Rest (Slipgate Studios, 2017) was used as stimuli C1 and C2 in the third and eighth sessions. The experience starts as the participants wake up in a hospital room environment, seeing their lower bodies, legs and feet on a hospital bed, unable to move their hands. Then a dark janitor starts to push the bed along the hospital halls as several jump scare events occur related to some phobias. Users cannot control the direction or speed of the translational movement, which is always forwards with smooth turns on the corners. The spatial scene changes as the hospital bed moves forward and users rotate their heads. The bed is probably working as a “real-world stabilized cue” (Jerald, 2015: 207-208) and reduces the cybersickness. Auditory cues are used to make users turn their heads to see some jump scare events, but those events mostly occur in the field of view of the users. Due to the slow speed of translational movement, the sense of vection is not very strong.

The Night the Carsons Disappeared (Long, 2016b), employed as stimulus D at fourth session, lets users experience a storyline about Michael Carson and her two daughters disappearing mysteriously from their home. Experience starts with a 30 s walking from street to a house and movement is not controlled by the users. Entering the house, users find themselves seated on an armchair, but they cannot see their body representation as an avatar. During the rest of the experience, users are directed to make rotational movements via auditory cues to observe objects supernaturally moving around the room. The only jump scare event is a spring toy box and the rest of the experience is mainly based on suspense of horror.

Affected: The Manor (Fallen Planet Studios, n. d.), which was the stimulus E in fifth session, requires a joystick for translational navigation in a haunted mansion. Rotational movements are also required to change the direction of translational movement, that is users walk along the center of the field of view. Although the second joystick arm can be used to control the direction of movement, participants were asked not to use this control during the experiment. Routes to four alternative endings are determined by the users, requiring them to make rotational movements in order to explore alternative routes in a maze of halls and galleries where they experience flashing lights and jump scare events such as bats, floating furniture and supernatural creatures. There is no representation of user as an avatar.

The Visitor (Long, 2016a), used as stimulus F in sixth session, utilizes a dark atmosphere to make users nervous and comfortable, rather than jump scare events. Visual cues such as flickering lights and shadows direct users' rotational movements. Users’ bodies or body parts are not represented. The environment is quite similar to stimulus A, in which users lie in a large bed.

Stimulus G employed in seventh session, Insane Decay of Mind: The Labyrinth (IV Productions, 2010), has a similar interaction mechanic to stimulus E. In addition, there are puzzles to be solved using gamepad controls. The environment is a school building that has wider and longer corridors compared to the halls in stimulus E. While exploring these corridors for puzzles, users need to avoid some of the non-player characters. Users’ bodies or body parts are not represented. There is small dot at the center of the viewpoint but it does not work as a “real-world stabilized cue”.

### 3.3. Procedure

The 16 item original English version of SSQ (Lane and Kennedy, 1988) was used for data collection with a four-point scale of 0 (none), 1 (slight), 2 (moderate) and 3 (severe). CSQ and VRSQ employ subsets of these original 16 items and F-SSQ operates a different factor structure based on the same items. The item set was administered immediately after each session. 9 sessions for each participant were spread to four weeks. Participants were exposed to the same stimuli at first and last sessions as “A1 - no experience” and “A2 - maximum experience” conditions, and at third and eighth sessions as “C1 - low experience” and “C2 - higher experience” conditions.

Participants were discouraged to attend multiple sessions on the same day. If necessary, there were at least 2 h between sessions and only two sessions were allowed per day. This approach is arguable since the sickness symptoms are reported to be outlasting after several hours following the experience (Gavgani et al., 2017). For this reason, we checked our dataset to identify the participants who engaged in two VR sessions on the same day. We detected that one participant took the session B after A1, three participants took the session C1 after B, five participants took D after C1, four participants took E after D, five participants took F after E, twelve participants took G after F, nine participants took C2 after G and another twelve participants took A2 after C2. We made a between-subjects comparison on mean scores of each symptom on last three groups. There was not a significant difference on any of the symptoms between the subjects who took G after F and who did not, although the scores were slightly higher for the ones who took two sessions on the same day. We also did not detect any significant differences between the participants who did and did not take the session C2 and A2 on the same day, while mean scores are similar on each symptom. Between the 9 participants who took C2 after G and 23 who did not; we detected a significant difference on two symptoms: “Blurred Vision” and “Dizziness with eyes wide open”. Surprisingly, these scores were lower for the participants who took the two sessions on the same day. We suppose that difference was due to interpersonal differences on participants’ vulnerability to cybersickness effects. The results are given on Table 3. Based on these results, we decided that after-effects of previous session did not reveal a significant effect on the session executed on the same day. In addition, it should be noted that SSQ based cybersickness scores were reported to decrease to pre-exposure levels after 30 min following the exposure, when the stimulus was not evoking high levels of cybersickness (Dziuda et al., 2014). As we further investigate in Results section, stimuli C and A and F did not evoke high scores compared to other stimuli. Stimuli G evoked the higher scores, but its after-effects did not yield to a significant effect on 14 symptoms, except” Blurred Vision” and “Dizziness with eyes open”, which were surprisingly lower for the group who attended two experiments on the same day. We did not investigate the other same-day sessions pairs as the occurrences of same-day sessions are relatively small in number.

### 3.4. Data analysis

Scales were scored using the methods suggested in related studies as explained under the heading “Psychometric Qualifications of SSQ and its variants”.

For each variant of SSQ, a CFA was conducted using Ωnyx structural equation modeling software (von Oertzen et al., 2015) to evaluate the construct validity of measurement models. The root mean square error of approximation (RMSEA), the root mean square residual (SRMR), the comparative fit index (CFI) and the Tucker–Lewis index (TLI) were calculated for each SSQ variant to verify the models’ fit to the data collected. Akaike information criterion (AIC) and the Bayes information criterion (BIC) were compared, as the smaller values indicate a better fit

**Table 3**
Item based between subjects score comparisons of sessions conducted on the same day and different days using the same stimuli.

| | | M | SD | T-test | | M | SD | T-test | | M | SD | T-test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| General discomfort | C2 after G | .11 | .33 | t(30) = -0.79, p > .05 | G after F | .75 | 1.14 | t(30) = 0.27, p > .05 | A2 after C2 | .42 | .67 | t(30) = -0.36, p > .05 |
| | C2 only | .30 | .70 | | G only | .65 | 0.93 | | A2 only | .50 | .61 | |
| Fatigue | C2 after G | .11 | .33 | t(30) = -0.43, p > .05 | G after F | 1.00 | 1.21 | t(30) = 0.99, p > .05 | A2 after C2 | .08 | .29 | t(30) = -0.86, p > .05 |
| | C2 only | .17 | .39 | | G only | .60 | 1.05 | | A2 only | .20 | .41 | |
| Headache | C2 after G | .11 | .33 | t(30) = 0.2, p > .05 | G after F | 1.17 | 1.19 | t(30) = 0.86, p > .05 | A2 after C2 | .08 | .29 | t(30) = -1.3, p > .05 |
| | C2 only | .09 | .29 | | G only | .85 | 0.88 | | A2 only | .35 | .67 | |
| Eye strain | C2 after G | .22 | .44 | t(30) = 0.02, p > .05 | G after F | 1.00 | 1.13 | t(30) = 1.08, p > .05 | A2 after C2 | .25 | .62 | t(30) = 0, p > .05 |
| | C2 only | .22 | .52 | | G only | .60 | 0.94 | | A2 only | .25 | .44 | |
| Difficulty focusing | C2 after G | .11 | .33 | t(30) = -0.14, p > .05 | G after F | .75 | 1.14 | t(30) = 0.72, p > .05 | A2 after C2 | .08 | .29 | t(30) = -0.54, p > .05 |
| | C2 only | .13 | .34 | | G only | .50 | 0.83 | | A2 only | .15 | .37 | |
| Salivation increasing | C2 after G | .11 | .33 | t(30) = 0.69, p > .05 | G after F | .25 | 0.62 | t(30) = 0, p > .05 | A2 after C2 | .17 | .39 | t(30) = 1.08, p > .05 |
| | C2 only | .04 | .21 | | G only | .25 | 0.44 | | A2 only | .05 | .22 | |
| Sweating | C2 after G | .11 | .33 | t(30) = -0.57, p > .05 | G after F | 1.00 | 0.95 | t(30) = 0.89, p > .05 | A2 after C2 | .25 | .62 | t(30) = 0.22, p > .05 |
| | C2 only | .22 | .52 | | G only | .65 | 1.14 | | A2 only | .20 | .62 | |
| Nausea | C2 after G | .11 | .33 | t(30) = -0.11, p > .05 | G after F | .50 | 0.90 | t(30) = -0.41, p > .05 | A2 after C2 | .08 | .29 | t(30) = 0.37, p > .05 |
| | C2 only | .13 | .46 | | G only | .65 | 1.04 | | A2 only | .05 | .22 | |
| Difficulty concentrating | C2 after G | .11 | .33 | t(30) = 0.2, p > .05 | G after F | .58 | 1.00 | t(30) = -0.04, p > .05 | A2 after C2 | .08 | .29 | t(30) = 0.37, p > .05 |
| | C2 only | .09 | .29 | | G only | .60 | 1.05 | | A2 only | .05 | .22 | |
| Fullness of the Head | C2 after G | .11 | .33 | t(30) = -0.14, p > .05 | G after F | .83 | 1.27 | t(30) = 0.56, p > .05 | A2 after C2 | .17 | .58 | t(30) = -0.41, p > .05 |
| | C2 only | .13 | .34 | | G only | .60 | 1.05 | | A2 only | .25 | .55 | |
| Blurred vision | C2 after G | 0.00 | .00 | t(22) = -2.15, p < .05 | G after F | .58 | 1.00 | t(30) = 1.03, p > .05 | A2 after C2 | .17 | .39 | t(30) = 0.54, p > .05 |
| | C2 only | .17 | .39 | | G only | .30 | 0.57 | | A2 only | .10 | .31 | |
| Dizziness with eyes open | C2 after G | 0.00 | .00 | t(22) = -2.47, p < .05 | G after F | .75 | 0.97 | t(30) = 0.58, p > .05 | A2 after C2 | .08 | .29 | t(30) = 1.31, p > .05 |
| | C2 only | .22 | .42 | | G only | .55 | 0.94 | | A2 only | 0.00 | .00 | |
| Dizziness with eyes closed | C2 after G | .11 | .33 | t(30) = 0.69, p > .05 | G after F | .67 | 0.98 | t(30) = 0.49, p > .05 | A2 after C2 | .08 | .29 | t(30) = 1.31, p > .05 |
| | C2 only | .04 | .21 | | G only | .50 | 0.89 | | A2 only | 0.00 | .00 | |
| Vertigo | C2 after G | .11 | .33 | t(30) = 0.69, p > .05 | G after F | .58 | 0.79 | t(30) = 1.31, p > .05 | A2 after C2 | .08 | .29 | t(30) = 1.31, p > .05 |
| | C2 only | .04 | .21 | | G only | .25 | 0.64 | | A2 only | 0.00 | .00 | |
| Stomach awareness | C2 after G | 0.00 | .00 | t(30) = -0.9, p > .05 | G after F | .33 | 0.89 | t(30) = -0.85, p > .05 | A2 after C2 | .08 | .29 | t(30) = 0.37, p > .05 |
| | C2 only | .09 | .29 | | G only | .65 | 1.09 | | A2 only | .05 | .22 | |
| Burping | C2 after G | 0.00 | .00 | t(30) = -0.62, p > .05 | G after F | 0.00 | 0.00 | t(30) = -1.09, p > .05 | A2 after C2 | .08 | .29 | t(30) = 1.31, p > .05 |
| | C2 only | .04 | .21 | | G only | .25 | 0.79 | | A2 only | 0.00 | .00 | |

of the model (Schreiber et al., 2006).

Discriminant validity was examined through Fornell-Larcker criterion for subdimensions of each scale. For each subdimension, the square root of average variance extracted (AVE) was checked to confirm its being greater than the correlations between the examined dimension and other dimensions of the same scale (Fornell and Larcker, 1981).

Reliability was explored in terms of internal consistency and test-retest reliability.

Test-retest reliability was indicated by Pearson correlations between the "A1 - no experience" and "A2 - maximum experience" conditions; and the "C1 - low experience" and "C2 - higher experience" conditions, for each scale and their subdimensions.

Internal consistency was indicated as Cronbach's alpha, which was calculated for all 288 cases in 9 sessions, and 224 cases in 7 sessions, excluding the no experience and low experience conditions explained above.

Sensitivity of the scales were explored through a series of ANOVAs for comparison of each stimulus' mean total severity score on each scale's subdimension scores, except CSQ which does not offer a total severity score. For CSQ, we provided the proportional subdimension score instead of total severity score, since it is suggested that proportionally higher Dizziness scores indicate higher levels of overall cybersickness. Proportional CSQ scores were adjusted using the method suggested by Drexler (2006) for SSQ subdimensions, in which the subdimension score was divided by the sum of all subdimensions. "A1 - No experience" and "C1 - low experience" conditions were excluded from the analysis of sensitivity. Scales' capability of providing significantly different scores for different stimuli is investigated as an indicator of sensitivity.

## 4. Results

### 4.1. Validity

As an indicator of construct validity, the RMSEA indicator of model fit was not below the acceptable value of < 0.06 for neither of the models, closest being 0.95 for CSQ. The SRMR indicator < 0.06 shows a model fit for VRSQ and CSQ, 0.037 and 0.038 consecutively. CFI > 0.95 for VRSQ and CSQ also indicate that these two models fit our data. None of the models fit our data based on TLI > 0.95 criteria. For these criteria of model fit, VRSQ and CSQ fit to our data according to SRMR and CFI indicators.

AIC and BIC were lowest for the CSQ, indicating its being the best fit of model to the data relative to the other three scales. Model fit indicators are presented in Table 4.

Exploring the discriminant validity of the subdimensions for each scale, our results showed that SSQ-O had a high correlation with SSQ-D ($r = .916$, $p > .01$) which was greater than square root of AVE (0.883) for SSQ-O and AVE for SSQ-D (0.897). All the other subdimensions

**Table 4**
Indicators of model fit for CFA on SSQ, F-SSQ, VRSQ and CSQ.

| | SSQ | F-SSQ | CSQ | VRSQ |
|---|---|---|---|---|
| $chi^2$ | 1149.608 | 612.722 | 92.694 | 97.393 |
| Restricted df | 97 | 104 | 26 | 26 |
| df | 105 | 120 | 36 | 36 |
| AIC | 7128.5 | 6879 | 3654 | 4294 |
| BIC | 7270.18 | 6996 | 3723 | 4363.5 |
| RMSEA (< .06) | 0.194 | 0.131 | 0.095 | 0.098 |
| SRMR (covariances only) (< .06 or < .08) | 0.381 | 0.19 | 0.038 | 0.037 |
| CFI (to independent model) > .95 | 0.625 | 0.825 | 0.951 | 0.954 |
| TLI (to independent model) > .95 | 0.594 | 0.798 | 0.932 | 0.936 |

**Table 5**
Indicators of Discriminant Validity. Values given in bold are √AVE. Italics are correlations ($p < .01$).

| | AVE | SSQ-N | SSQ-O | SSQ-D |
|---|---|---|---|---|
| SSQ-N | 0.802 | **0.896** | | |
| SSQ-O | 0.779 | *0.855* | **0.883** | |
| SSQ-D | 0.805 | *0.840* | *0.916* | **0.897** |
| | | FSSQ-N | FSSQ-O | |
| FSSQ-N | 0.806 | **0.898** | | |
| FSSQ-O | 0.934 | *0.854* | **0.967** | |
| | AVE | VRSQ-O | VRSQ-D | |
| VRSQ-O | 0.766 | **0.875** | | |
| VRSQ-D | 0.834 | *0.837* | **0.913** | |
| | AVE | CSQ Dizziness | CSQ Difficulty Focusing | |
| CSQ Dizziness | 0.808 | **0.899** | | |
| CSQ Difficulty Focusing | 0.765 | *0.781* | **0.875** | |

were confirmed to fulfill Fornell-Larcker criterion, indicating discriminant validity of the scales. Scale correlations and AVE values for subdimensions are given in Table 5.

### 4.2. Reliability

Reliability evaluation in terms of internal consistency provided evidence of reliability for all scales, based on Cronbach's alpha indicator. Test-retest reliability changes from insignificant or weak to moderate positive Pearson correlations between different levels of experience.

Test-Retest reliability indicated by Pearson r value was not significant for SSQ-O, SSQ-D, F-SSQ-O, VRSQ-O, VRSQ Total and CSQ Difficulty in Focusing scores of participants in “A1 - no experience” and “A2 - maximum experience” conditions ($p > .05$). The scores that significantly correlated between these conditions varied from weak to moderate, as can be followed on Table 6, which are given in bold. On the contrary, all the scale the scores in C1 - low/C2 - higher experience conditions correlated significantly at a moderate level.

The Cronbach alpha scores obtained through the data acquired in 9 sessions were above 0.7 for all scales and subdimensions, indicating an adequate internal consistency (see Table 6). When the data from no A1 - experience and C1 - low experience conditions were discarded, internal reliability increased slightly for all measures, as can be followed on Table 6.

**Table 6**
Indicators of reliability. Significant Pearson r values are given in bold. Significance levels can be followed on the table.

| | Internal Consistency | | Test-retest Reliability | | | |
|---|---|---|---|---|---|---|
| | 9 sessions | 7 sessions | No/Max. Exp. | | Low/Higher Exp. | |
| | Cronbach's α | Cronbach's α | r | Sig. (p) | r | Sig. (p) |
| SSQ N | .84 | .85 | **.45** | **.01** | **.64** | **.000** |
| SSQ O | .91 | .93 | .31 | .08 | **.57** | **.001** |
| SSQ D | .88 | .90 | .28 | .12 | **.73** | **.000** |
| SSQ TS | .94 | .95 | **.38** | **.03** | **.71** | **.000** |
| F-SSQ N | .86 | .88 | **.49** | **.01** | **.63** | **.000** |
| F-SSQ O | .91 | .93 | .35 | .053 | **.57** | **.001** |
| F-SSQ TS | .94 | .95 | **.39** | **.03** | **.72** | **.000** |
| VRSQ O | .87 | .90 | .24 | .18 | **.53** | **.002** |
| VRSQ D | .84 | .85 | **.44** | **.01** | **.73** | **.000** |
| VRSQ Total | .92 | .93 | .30 | .1 | **.67** | **.000** |
| CSQ Dizziness | .84 | .85 | **.37** | **.04** | **.58** | **.000** |
| CSQ Diff. in Foc. | .84 | .85 | .16 | .39 | **.49** | **.005** |

### 4.3. Sensitivity

#### 4.3.1. Total severity scores

As determined by one-way ANOVAs for each scale's TS (total severity) score, there was a significant effect of the stimuli on mean SSQ TS score ($F(6,217) = 5.68$, $p < .05$), F-SSQ TS score ($F(6,217) = 5.99$, $p < .05$) and VRSQ TS score ($F(6,217) = 5.43$, $p < .05$).

According to Tukey post hoc tests, SSQ TS score was significantly higher for stimulus “E” compared to A-2, C-2, F and B. Stimulus G score was significantly higher than stimuli F, A2 and C2. Stimulus D scores were higher than the B, F, A2 and C2 scores but differences were not statistically significant. Stimulus B, F, C2, and A2 had closer magnitudes and their differences were not statistically significant.

The F-SSQ TS scores were significantly higher for stimulus E compared to C-2, A-2, F and B, same as the SSQ-TS. Also, stimulus “G” F-SSQ TS score was significantly higher than stimuli C, C2 and A2. Score for stimulus D is higher than the scores of B, F, A2 and C2 but mean differences were not significant, as observed in SSQ TS scores. Similarly, stimulus “B, F, C2, and A2” are closer to each other without any significant differences.

Exploring the VRSQ TS scores through Tukey post hoc tests, we determined that stimulus E score was significantly higher than F, A2 and C2. Stimulus G score was significantly higher than stimuli A2 and C2. Unlike SSQ and F-SSQ, VRSQ TS score for stimulus D is significantly higher than stimulus C2. D score is also higher than B, F and A2, but the mean differences were not significant. Score magnitudes were closer for the other stimuli, as observed in SSQ and F-SSQ total severity scores. Mean differences between the stimuli can be seen in Table 7 where significant differences ($p < .05$) are given in bold.

There is not a method for calculating a total severity score for CSQ, but Stone III (2017) suggested that proportionally higher Dizziness mean scores indicate a higher level of cybersickness. As seen on Fig. 2, proportional Dizziness score is higher than Difficulty Focusing for stimuli E and G. This may suggest that CSQ confirms with the results of other scales' TS scores, as E and G yielded highest total severity scores on SSQ, FSSQ and VRSQ.

#### 4.3.2. Subdimension scores

A series of one-way ANOVAs revealed significant effects of different stimuli on mean subdimension scores.

There is a significant effect of different VEs on SSQ-N ($F(6,217) = 4.95$, $p < .05$), SSQ-O ($F(6,217) = 4.78$, $p < .05$) and SSQ-D ($F(6,217) = 6.4$, $p < .05$) scores.

The significant differences between stimuli mean scores according to Tukey post hoc tests are given in Table 8, along with the stimuli ordered by subdimension scores.

The effect on SSQ-N score was due to significant differences of stimuli E and G scores being higher than three of other stimuli, F, C2 and A2. On SSQ-O, E and G scores differentiated significantly from two other stimuli, C2 and A2. SSQ-D score of stimuli E and G were significantly higher than other four, B, F, C2 and A2. Results suggested that SSQ-D is capable of distinguishing the differences between applications as well as SSQ-N and SSQ-O.

We observed a significant effect of different VEs on F-SSQ-N ($F(6,217) = 6.9$, $p < .05$) and F-SSQ-O ($F(6,217) = 4.95$, $p < .05$). F-SSQ N provided similar scores with SSQ-N in terms of distinguishing software, with an addition that detecting a significant difference between E and B, which was not detected by SSQ-N. FSSQ-O and SSQ-O yielded same results in terms sensitivity. FSSQ had more items per subdimension that it was expected to provide higher sensitivity.

VRSQ-D ($F(6,217) = 6.33$, $p < .05$) and VRSQ-O ($F(6,217) = 4.17$, $p < .05$) were also affected by different stimuli. VRSQ-D scores showed

**Table 7**
Mean Differences of Total Severity Scores. Mean differences given in bold are significant at 0.05 level.

| SSQ Total Severity Mean Differences | | | | | | | |
|---|---|---|---|---|---|---|---|
| | E | G | D | B | F | A2 | C2 |
| E | | 3.74 | 13.56 | **30.62** | **34.83** | **37.87** | **39.39** |
| G | | | 9.82 | 26.88 | **31.09** | **34.13** | **35.65** |
| D | | | | 17.06 | 21.27 | 24.31 | 25.83 |
| B | | | | | 4.21 | 7.25 | 8.77 |
| F | | | | | | 3.04 | 4.56 |
| A2 | | | | | | | 1.52 |
| Mean | 49.32 | 45.58 | 35.76 | 18.70 | 14.49 | 11.45 | 9.93 |
| St.Dev. | 59.95 | 55.80 | 48.76 | 21.93 | 26.68 | 19.00 | 19.67 |
| **FSSQ Total Severity Mean Differences** | | | | | | | |
| | E | G | D | B | F | A2 | C2 |
| E | | .07 | .34 | **.80** | **.89** | **.97** | **1.01** |
| G | | | .27 | **.74** | **.82** | **.90** | **.94** |
| D | | | | .46 | .55 | .63 | .67 |
| B | | | | | .09 | .17 | .20 |
| F | | | | | | .08 | .12 |
| A2 | | | | | | | .04 |
| Mean | 1.25 | 1.19 | .91 | .45 | .36 | .28 | .25 |
| St.Dev. | 1.51 | 1.42 | 1.24 | .57 | .70 | .49 | .48 |
| **VRSQ Total Severity Mean Differences** | | | | | | | |
| | E | G | D | B | F | A2 | C2 |
| E | | 2.14 | 4.30 | 14.51 | **16.33** | **17.37** | **19.04** |
| G | | | 2.16 | 12.37 | 14.19 | **15.23** | **16.90** |
| D | | | | 10.21 | 12.03 | 13.07 | **14.74** |
| B | | | | | 1.82 | 2.86 | 4.53 |
| F | | | | | | 1.04 | 2.71 |
| A2 | | | | | | | 1.67 |
| Mean | 23.72 | 21.59 | 19.43 | 9.22 | 7.40 | 6.35 | 4.69 |
| St.Dev. | 28.03 | 26.55 | 25.07 | 11.66 | 15.08 | 9.96 | 9.46 |

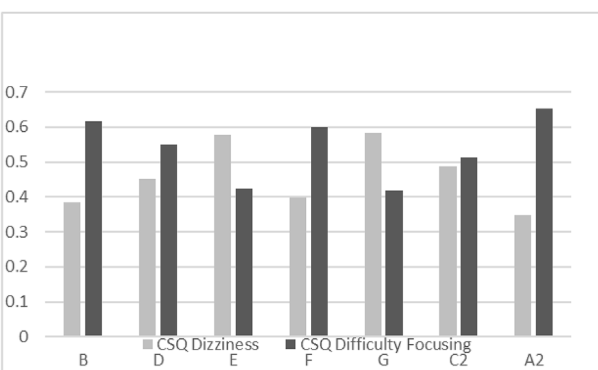


**Fig. 2.** CSQ proportional scores.

that both E and G scores were significantly higher than four other stimuli, B, F, C2 and A2. Besides, it detected a significant difference between D and C2. For VRSQ-O, only stimulus E score was found to be higher than F, C2 and A2. The difference of E from F was not detected by SSQ-O and FSSQ-O, but SSQ-O yielded that G is significantly higher than A2 and C2, which could not be detected by VRSQ-O.

There was a significant effect of different VEs on CSQ Dizziness ($F(6,217) = 8.76$, $p < .05$) and Difficulty Focusing ($F(6,217) = 4.18$, $p < .005$) dimensions. CSQ Dizziness score yielded the same results with VRSQ-D for significant differences of E and G. However it could not detect a significant difference between D and C2. Difficulty Focusing score was significant for D and E, significantly higher than C2 and A2.

**Table 8**
Subdimension score magnitudes and significant differences between scores.

| | Order of Score Magnitudes | Significant Differences |
|---|---|---|
| SSQ N | E > G > D > B > A2 > F > C2 | E > F, C2, A2<br>G > F, C2, A2 |
| SSQ O | E > G > D > B > F > A2 > C2 | E > C2, A2<br>G > C2, A2 |
| SSQ D | E > G > D > B > F > C2 > A2 | E > B, F, C2, A2<br>G > B, F, C2, A2 |
| VRSQ O | E > G > D > B > F = A2 > C2 | E > F, C2, A2 |
| VRSQ D | E > G > D > B=F > A2 > C2 | D > C2<br>E > B, F, C2, A2<br>G > B, F, C2, A2 |
| FSSQ N | E > G > D > B > A2 > C2=F | E > B, F, C2, A2<br>G > F, C2, A2 |
| FSSQ O | G > E > D > B > F > A2 > C2 | E > C2, A2<br>G > C2, A2 |
| CSQ Dizziness | E > G > D > F > B > C2 > A2 | E > B, F, C2, A2<br>G > B, F, C2, A2 |
| CSQ Diff. Focusing | E > D > G > B > F > A2 > C2 | D > C2, A2<br>E > C2, A2 |

## 5. Discussion

Results based on CFA did not suggest evidence of construct validity for SSQ and FSSQ in evaluation of cybersickness experienced using publicly available VE software engaged through a consumer HMD. On the other hand, there was evidence of construct validity for VRSQ and CSQ. CSQ had been developed based on a dataset collected through the publicly available VE software, similar to our dataset, while VRSQ had been developed using a dataset collected through a custom target-selection application. For this reason, CSQ might have showed a slightly better fit to our data, compared to VRSQ. Although VRSQ and CSQ have 5 symptoms in common among their 9 items, their reduced item structure represented the dimensions of cybersickness more accurately, compared to 16 item SSQ and F-SSQ.

The high correlations within the scales' subdimensions imply multicollinearity. As the scoring method of SSQ employs several symptoms in multiple dimensions, that was not unexpected for SSQ subdimensions. Based on Fornell-Larcker criterion, we verified that SSQ-O and SSQ-D dimensions did not provide discriminant validity for evaluation of consumer VEs. We provided evidence for discriminant validity of constructs represented in other scales, suggesting that cybersickness is more likely to be a bi-dimensional construct.

We provided evidence on reliability of SSQ and its variants. However, in terms of test-retest reliability, the self-reported symptoms of cybersickness did not correlate sufficiently in repeated experiments. The correlation between test/re-test scores were affected by the experience level of users. The self-reported symptoms of cybersickness changed as users engaged more with different VEs. This might have been due to the users' adaptation to the VR. On the other hand, the weaker and insignificant correlation of very first engagement with the repeated measurement might be due to the self-induced motion sickness (Almeida et al., 2017), that occurs when the VR users are informed about the side effects of HMD use before they experience it (Young et al, 2007). However, the moderate correlations between low and higher experience conditions also suggest that cybersickness measures are not strongly reliable when the users' levels of experience vary. Still, reliability by means of internal consistency is adequate when lowest experience cases are included in analysis.

When the sensitivity of TS (total severity) scores are inspected, VRSQ TS scoring method is less capable of distinguishing the level of cybersickness based on the application aspects, compared to SSQ and FSSQ, since VRSQ has fewer items. Although CSQ does not provide a total severity measure, the CSQ profiles of stimuli that cause a higher level of cybersickness are different from the profiles of other stimuli.

The profile based on proportional CSQ scores yields a higher Dizziness score when stimulus triggers vection, which is known to be an application aspect that causes motion sickness. CSQ proportional Dizziness score being higher than Difficulty Focusing score indicated higher levels of cybersickness, as suggested by its developer.

When the subdimension scores were evaluated, VRSQ-D and CSQ Dizziness scores were found to be highly sensitive to application aspects, compared to SSQ-D. Absence of Nausea related items; burping, stomach awareness, increased salivation and sweating; did not decrease the sensitivity of subdimensions in CSQ and VRSQ and it was possible to compare different VEs without the nausea dimension. However, it should be noted that the SSQ-N scores were also highly sensitive to differences between VEs. Besides the SSQ-N Nausea symptom is still a part of CSQ Dizziness and SSQ-N General Discomfort symptom is included in VRSQ-D. Furthermore, sweating cannot be evaluated by participants as a self-report measure, although it was found to be "the best physiological correlate of motion sickness" when it is measured through forehead skin conductance (Gavgani et al., 2017). This kind of sweating can be worded as "cold sweating" as it is referred in MSAQ (Gianaros et al., 2001), which is different from sweating due to physical effort and arousal in relation with the virtual experience. However, sweating "mainly occurs during the presence of the provocative stimulus and quickly dissipates upon its termination" (Gavgani et al., 2017) and due to the post-experiment use of questionnaires, participants are likely to fail in reporting the sweating symptom. Regarding to this finding, we suggest that the "Increased Salivation" symptom should also be explored comparing the biometric observations and self-reports, to identify either it does not occur in case of cybersickness or it cannot be reported by the participants after the experiment.

Based on the literature on "Application Aspects Affecting Cybersickness", stimuli E, G and C were suspected to trigger higher levels of cybersickness, as they allow translational navigation as well as rotational navigation, and translational navigation which yields vection. Stimulus C was expected to cause less cybersickness compared to E and G, since the translational movement is only in fore-axis with a constant speed, not controlled by the user. Stimuli A, B, D and F only allow rotational navigation, where users are located in virtual rooms in which only certain objects move. Usually, users are stimulated by directional audio to rotate their heads to a direction that they would see a moving object. The longer session durations are likely to cause higher levels of cybersickness in these type of VEs, instead of vection.

There was evidence of sensitivity to application aspects, but the results of stimuli C, E and G triggering the highest level of cybersickness based on vection due to discordant translational movement were not as expected. For stimuli E, it should be also considered that the experiment duration, which around 11 min in average, was another reason that triggers cybersickness, as Min et al. (2004) reported that SSQ scores significantly increase after 10 min of simulation experience. On the other, we did not observe a high level of sickness on stimuli D, which does not involve any translational movement, although participants were immersed into it for a longer duration. Our results suggest that cybersickness is not only "a function of the time spent in simulator", but mainly it is highly related with vection. The relationship of vection and time spent in the simulator as a cause of cybersickness should be explored further.

Although the highest scores for all measures were on stimuli E and G, stimuli B and D provided higher scores than C for most of the measures. As applications B and D require a lot of rotational movements to follow the objects in the scene, this may have also caused cybersickness. The slow-pace automated translational movement in stimuli C did cause a lower level cybersickness than we expected.

## 6. Conclusion

Our results suggested that CSQ and VRSQ, which were specifically designed for measuring cybersickness, had better psychometric qualities for assessing HMD VR applications, when they were compared to SSQ and F-SSQ, which were tools that were intended to measure simulator sickness. We provided evidence for validity of CSQ and VRSQ as cybersickness measures, while SSQ and F-SSQ could not be validated psychometrically. Our results provided evidence for reliability of all measures. VRSQ and CSQ were highly sensitive to differences between the application aspects of evaluated VEs, although they investigate fewer symptoms than simulator sickness scales.

However, since the psychometric qualities of CSQ and VRSQ are quite similar, trying to decide which one is superior to the other in evaluation of cybersickness in consumer VEs becomes difficult.

When we investigated the symptoms employed in these scales, there were similarities as well as inconsistencies.

The three symptoms that were mainly related to digestive system; Increased Salivation, Burping and Stomach Awareness; were excluded both in CSQ and VRSQ, but CSQ still incorporated the Nausea symptom as an indicator of cybersickness, which is also related to the digestive system.

Eye Strain, Difficulty in Focusing and Blurred Vision are used in both scales as ocular symptoms of cybersickness. However, VRSQ evaluates the General Discomfort and Fatigue items in the same dimension with two of those ocular system symptoms, while it uses Blurred Vision as an indicator of Disorientation. Other two common items, Dizziness (eyes closed) and Vertigo are symptoms mainly related to vestibular system. The Headache and Fullness of Head symptoms are effects that can be related to nervous system. While CSQ operates Fullness of Head among with symptoms related with ocular system, VRSQ uses it along with items related to vestibular system. In both scales, Headache is evaluated among the vestibular system symptoms.

We should not disregard the fact that CSQ was developed through an item-response theory approach, while VRSQ developers followed a factor analysis based method. VRSQ was based on a sample of 16 different users, while CSQ was developed on data collected from 202 different participants. The stimuli in the experiments of VRSQ was limited to four different selection tasks with mobile HMDs, as CSQ participants evaluated one of the two consumer VEs that had different navigation mechanisms. CSQ uses a complicated scoring system based on item weights. Considering these limitations in their development process, neither CSQ nor VRSQ have evolved in ideal conditions that represent the applications and users of consumer VEs.

Our study employed a broader range of consumer VE applications with different methods of interaction and navigation, compared to both VRSQ and CSQ studies. On the other hand, our participants were limited to 32 individuals, which cannot be considered as a representation of the population of VR consumers.

The factor analytic approach in development on SSQ should be considered carefully, as its succeeding variants followed the same approach in determining symptoms and dimensions of cybersickness. Some symptoms that were not consistently reported by participants were eliminated during the development of these scales, as a characteristic of scale development methodology. However, the eliminated symptoms might not have been evoked in participants due to the limited characteristics of stimuli employed in the study or limited sample characteristics of participants. Researchers also consider the findings of recent research (Gavgani et al., 2018) that claims cybersickness symptoms are not different from motion sickness symptoms. Thus, SSQ, which excluded sopite symptoms of Pensacola Motion Sickness Questionnaire, may not be correctly including relevant symptoms of cybersickness at all.

These limitations prevented us from offering a new scale based on SSQ items. Although our results also supported CSQ and VRSQ for their assumption that the symptoms in digestive system are not reported in case of cybersickness, we suggest further research in order to cover wider user populations and different stimuli that represent broader application aspects of consumer VEs. For this reason, we decided to publish our dataset as a supplementary material along with our work, and ask other researchers act in the same way.

## Declarations of interest

None.

## References


Almeida, A., Rebelo, F., Noriega, P., Vilar, E., 2017. July). Virtual reality self induced cybersickness: an exploratory study. In: International Conference on Applied Human Factors and Ergonomics. Springer, Cham, pp. 26–33. https://doi.org/10.1007/978-3-319-60582-1_3.

Alshaer, A., Regenbrecht, H., O'Hare, D., 2017. Immersion factors affecting perception and behaviour in a virtual reality power wheelchair simulator. Appl. Ergon. 58, 1–12. https://doi.org/10.1016/j.apergo.2016.05.003.

Ames, S.L., Wolffsohn, J.S., Mcbrien, N.A., 2005. The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. Optom. Vis. Sci. 82 (3), 168–176. https://doi.org/10.1097/01.OPX.0000156307.95086.6.

Aykent, B., Yang, Z., Merienne, F., Kemeny, A., 2014, September. Simulation sickness comparison between a limited field of view virtual reality head mounted display (Oculus) and a medium range field of view static ecological driving simulator (Eco2). In: Driving Simulation Conference Europe 2014 Proceedings. Society for Modeling & Simulation International, pp. 65–71.

Bonato, F., Bubka, A., Palmisano, S., 2009. Combined pitch and roll and cybersickness in a virtual environment. Aviat. Space Environ. Med. 80 (11), 941–945. https://doi.org/10.3357/ASEM.2394.2009.

Bonato, F., Bubka, A., Palmisano, S., Phillip, D., Moreno, G., 2008. Vection change exacerbates simulator sickness in virtual environments. Presence Teleoperators Virtual Environ. 17 (3), 283–292. https://doi.org/10.1162/pres.17.3.283.

Bouchard, S., Robillard, G., Renaud, P., 2007. Revising the factor structure of the simulator sickness questionnaire. Summer. Annu. Rev. Cybertherapy Telemed. 5, 128–137.

Brooks, J.O., Goodenough, R.R., Crisler, M.C., Klein, N.D., Alley, R.L., Koon, B.L., et al., 2010. Simulator sickness during driving simulation studies. Accid. Anal. Prev. 42, 788–796.

Bruck, S., Watters, P.A., 2011. The factor structure of cybersickness. Displays 32 (4), 153–158. https://doi.org/10.1016/j.displa.2011.07.002.

Chan, C.L.F., Ngai, E.K.Y., Leung, P.K.H., Wong, S., 2010. Effect of the adapted virtual reality cognitive training program among Chinese older adults with chronic schizophrenia: a pilot study. Int. J. Geriatr. Psychiatry 25 (6), 643–649. https://doi.org/10.1002/gps.2403.

Chardonnet, J.R., Mirzaei, M.A., Merienne, F., 2017. Features of the postural sway signal as indicators to estimate and predict visually induced motion sickness in virtual reality. Int. J. Hum. Comput. Interact. 33 (10), 771–785. https://doi.org/10.1080/10447318.2017.1286767.

Chen, W., Plancoulaine, A., Férey, N., Touraine, D., Nelson, J., Bourdot, P., 2013. October). 6DoF navigation in virtual worlds: comparison of joystick-based and head-controlled paradigms. ACM In: Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology, pp. 111–114. https://doi.org/10.1145/2503713.2503754.

Cobb, S.V.G., 1999. Measurement of postural stability before and after immersion in a virtual environment. Appl. Ergon. 30 (1), 47–57. https://doi.org/10.1016/S0003-6870(98)00038-6.

Cobb, S.V., Nichols, S., Ramsey, A., Wilson, J.R., 1999. Virtual reality-induced symptoms and effects (VRISE). *Presence*. Teleoperators Virtual Environ. 8 (2), 169–186. https://doi.org/10.1162/105474699566152.

Dennison, M.S., D'Zmura, M., 2017. Cybersickness without the wobble: experimental results speak against postural instability theory. Appl. Ergon. 58, 215–223. https://doi.org/10.1016/j.apergo.2016.06.014.

Dorado, J.L., Figueroa, P.A., 2014, March. Ramps are better than stairs to reduce cybersickness in applications based on a HMD and a gamepad. In: 3D User Interfaces (3DUI), 2014 IEEE Symposium on, pp. 47–50. https://doi.org/10.1109/3DUI.2014.6798841. IEEE.

Draper, M.H., Viirre, E.S., Furness, T.A., Gawron, V.J., 2001. Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. Hum. Factors 43 (1), 129–146. https://doi.org/10.1177/154193120004400512.

Drexler, J., 2006. Identification of System Design Features that Affect Sickness in Virtual Environments. Unpublished doctoral dissertation University of Central Florida, Orlando 2006.

Dziuda, Ł., Biernacki, M.P., Baran, P.M., Truszczyński, O.E., 2014. The effects of simulated fog and motion on simulator sickness in a driving simulator and the duration of after-effects. Appl. Ergon. 45 (3), 406–412. https://doi.org/10.1016/j.apergo.2013.05.003.

Epic Games, 2015. Virtual Reality Best Practices in Unreal Engine 4 Documentation. Retrieved from. https://docs.unrealengine.com/en-us/Platforms/VR/ContentSetup#vrandsimulationsickness.

Fallen Planet Studios Affected - the Manor. n.d, [Store Page]. Retrieved from. https://www.oculus.com/experiences/gear-vr/1058194390920956.

Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. J. Mark. Res. 39–50. https://doi.org/10.2307/3151312.

Garris-Reif, R., Franz, T.M., 1995. Simulator sickness and human task performance in conventional and virtual environments. In: Advances in Industrial Ergonomics and Safety VII. Taylor & Francis.

Gavgani, A.M., Nesbitt, K.V., Blackmore, K.L., Nalivaiko, E., 2017. Profiling subjective symptoms and autonomic changes associated with cybersickness. Auton. Neurosci.: Basic Clin. 203, 41–50. https://doi.org/10.1016/j.autneu.2016.12.004.

Gavgani, A.M., Hodgson, D.M., Nalivaiko, E., 2017. Effects of visual flow direction on signs and symptoms of cybersickness. PLoS One 12 (8), e0182790.

Gavgani, A.M., Walker, F.R., Hodgson, D.M., Nalivaiko, E., 2018. A comparative study of cybersickness during exposure to virtual reality and "classic" motion sickness: are they different? J. Appl. Physiol. 125 (6), 1670–1680.

Gianaros, P.J., Muth, E.R., Mordkoff, J.T., Levine, M.E., Stern, R.M., 2001. A questionnaire for the assessment of the multiple dimensions of motion sickness. Aviat. Space Environ. Med. 72 (2), 115–119.

Google Designing for google cardboard – physiological considerations. n.d, Retrieved from. https://designguidelines.withgoogle.com/cardboard/designing-for-google-cardboard/physiological-considerations.html, Accessed date: 10 March 2018.

Graybiel, A., Knepton, J., 1976. Sopite syndrome: a sometimes sole manifestation of motion sickness. Aviat. Space Environ. Med. 47 (8), 873–882.

Graybiel, A., Wood, C.D., Miller, E.F., Cramer, D.B., 1968. Diagnostic criteria for grading the severity of acute motion sickness. Aero. Med. 39 (5), 453–455.

Greenlight, V.R., December 3 2015. Virtual Reality Interest Highest Among Gen Z. VR Even Appeals to Baby Boomers. https://www.emarketer.com/Article/Virtual-Reality-Interest-Highest-Among-Gen-Z/1013295.

Hakkinen, J., Vuori, T., Paakka, M., 2002, October. Postural stability and sickness symptoms after HMD use. IEEE Int. Conf. Syst. Man Cybern. 1, 147–152.

Hammerhead, V.R., 2016, June 23. ABE VR (VR Experience). Retrieved from. https://store.steampowered.com/app/458590/ABE_VR/.

Helland, A., Lydersen, S., Lervåg, L.E., Jenssen, G.D., Mørland, J., Slørdal, L., 2016. Driving simulator sickness: impact on driving performance, influence of blood alcohol concentration, and effect of repeated simulator exposures. Accid. Anal. Prev. 94, 180–187.

Hettinger, L.J., Riccio, G.E., 1992. Visually induced motion sickness in virtual environments. Presence: Teleoperators & Virtual Environments 1 (3), 306–310. https://doi.org/10.1162/pres.1992.1.3.306.

Howarth, P.A., Finch, M., 1999. The nauseogenicity of two methods of navigating within a virtual environment. Appl. Ergon. 30 (1), 39–45. https://doi.org/10.1016/S0003-6870(98)00041-6.

IV Productions, 2010, May 6. Insane Decay of Mind: the Labyrinth. [Store page]. Retrieved from. https://store.steampowered.com/app/450060/Insane_Decay_of_Mind_The_Labyrinth/.

Iwata, H., 2013. Locomotion interfaces. In: Steinicke, F., Visell, Y., Campos, J., Lécuyer, A. (Eds.), Human Walking in Virtual Environments. Springer, New York, NY, pp. 199–219. https://doi.org/10.1007/978-1-4419-8432-6_9.

Jerald, J., 2015. The VR Book: Human-Centered Design for Virtual Reality. ÜLKE. Association for Computing Machinery - Morgan & Claypoolhttps://doi.org/10.1145/1792790.2792811.

Kellogg, R.S., Kennedy, R.S., Graybiel, A., 1965. Motion sickness symptomatology of labyrinthine defective and normal subjects during zero gravity maneuvers. Aerosp. Med. 36 (1965), 315–318.

Kennedy, R.S., Drexler, J., Kennedy, R.C., 2010. Research in visually induced motion sickness. Appl. Ergon. 41 (4), 494–503. https://doi.org/10.1016/j.apergo.2009.11.006.

Kennedy, R.S., Jones, M.B., Lilienthal, M.G., Harm, D.L., 1994, May. Profile analysis of after-effects experienced during exposure to several virtual reality environments. In: AGARD Conference Proceedings, 2-2.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.E., 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. The International Journal of Aviation Psychology 3 (3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3.

Kennedy, R.S., Lane, N.E., Lilienthal, M.G., Berbaum, K.S., Hettinger, L.J., 1992. Profile analysis of simulator sickness symptoms: application to virtual environment systems. Presence Teleoperators Virtual Environ. 1 (3), 295–301. https://doi.org/10.1162/pres.1992.1.3.295.

Kennedy, R.S., Lilienthal, M.G., Berbaum, K.S., Baltzley, D.R., McCauley, M.E., 1989. Simulator sickness in US Navy flight simulators. Aviat. Space Environ. Med. 60 (1), 10–16.

Kennedy, R.S., Graybiel, A., 1965. The Dial Test: a A Standardized Procedure for the Experimental Production of Canal Sickness Symptomatology in a Rotating Environment (Rep. No. 113, NSAM 930) Naval School of Aerospace Medicine. (Pensacola, FL).

Keshavarz, B., Hecht, H., 2011a. Axis rotation and visually induced motion sickness: the role of combined roll, pitch, and yaw motion. Aviat. Space Environ. Med. 82 (11), 1023–1029. https://doi.org/10.3357/ASEM.3078.2011.

Keshavarz, B., Hecht, H., 2011b. Validating an efficient method to quantify motion sickness. Hum. Factors 53 (4), 415–426. https://doi.org/10.1177/0018720811403736.

Keshavarz, B., Hecht, H., 2014. Pleasant music as a countermeasure against visually induced motion sickness. Appl. Ergon. 45 (3), 521–527. https://doi.org/10.1016/j.apergo.2013.07.009.

Kim, D.H., Parker, D.E., Park, M.Y., 2004. A New Procedure for Measuring Simulator Sickness–The RSSQ. Human Interface Technology Laboratory, University of Washington, Seattle, WA.

Kim, H.K., Park, J., Choi, Y., Choe, M., 2018. Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. Appl. Ergon. 69, 66–73. https://doi.org/10.1016/j.apergo.2017.12.016.

Lampton, D.R., Rodriguez, M.E., Cotton, J.E., 2000, July. Simulator sickness symptoms during team training in immersive virtual environments. No. 5 In: Proceedings of the

Human Factors and Ergonomics Society Annual Meeting 44. SAGE Publications, Sage CA: Los Angeles, CA, pp. 530–533.
Lane, N.E., Kennedy, R.S., 1988. A New Method for Quantifying Simulator Sickness: Development and Application of the Simulator Sickness Questionnaire (SSQ). Essex Corporation, Orlando, FL 88-7.
LaViola Jr., J.J., 2000. A discussion of cybersickness in virtual environments. ACM SIGCHI Bull. 32 (1), 47–56. https://doi.org/10.1145/333329.333344.
Lin, J.W., Duh, H.B.L., Parker, D.E., Abi-Rached, H., Furness, T.A., 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In: Virtual Reality, 2002. Proceedings. IEEE. IEEE, pp. 164–171.
Long, S., 2016a. The Visitor. https://store.steampowered.com/app/457580/The_Visitor/ [Store page] Retrieved from, April 1.
Long, S., 2016b. The Night the Carsons disappeared. [Store page] Retrieved from, October 24. https://store.steampowered.com/app/543350/The_Night_The_Carsons_Disappeared/.
McCauley, M., Hettinger, L., Sharkey, T., Sinacori, J., 1990. The effects of simulator visual-motion asynchrony on simulator induced sickness. In: Proceedings of the Flight Simulation Technologies Conference and Exhibit, Dayton, OH, https://doi.org/10.2514/6.1990-3172.
McCauley, M.E., Sharkey, T.J., 1992. Cybersickness: perception of self-motion in virtual environments. Presence Teleoperators Virtual Environ. 1 (3), 311–318. https://doi.org/10.1162/pres.1992.1.3.311.
Mgsstudio, 2016. DemonicGuestVR (VR experience). Retrieved from. https://store.steampowered.com/app/542820/DemonicGuestVR/ October 18.
Min, B.C., Chung, S.C., Min, Y.K., Sakamoto, K., 2004. Psychophysiological evaluation of simulator sickness evoked by a graphic simulator. Appl. Ergon. 35 (6), 549–556. https://doi.org/10.1016/j.apergo.2004.06.002.
Moss, J.D., Muth, E.R., 2011. Characteristics of head-mounted displays and their effects on simulator sickness. Hum. Factors 53 (3), 308–319. https://doi.org/10.1177/0018720811405196.
Nalivaiko, E., Davis, S.L., Blackmore, K.L., Vakulin, A., Nesbitt, K.V., 2015. Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. Physiol. Behav. 151, 583–590. https://doi.org/10.1109/SeGAH.2017.7939283.
Nelson, W.T., Roe, M.M., Bolia, R.S., Morley, R.M., 2000. Assessing Simulator Sickness in a See-Through HMD: Effects of Time Delay, Time on Task, and Task Complexity (No. ASC-00-1047). AIR FORCE RESEARCH LAB WRIGHT-PATTERSON AFB OH.
Nesbitt, K., Davis, S., Blackmore, K., Nalivaiko, E., 2017. Correlating reaction time and nausea measures with traditional measures of cybersickness. Displays 48, 1–8.
Nichols, S., Patel, H., 2002. Health and safety implications of virtual reality: a review of empirical evidence. Appl. Ergon. 33 (3), 251–271. https://doi.org/10.1016/S0003-6870(02)00020-0.
Oculus VR, LLC, 2017. Oculus Best Practices Version 310-30000-02. Retrieved from. https://static.oculus.com/documentation/pdfs/intro-vr/latest/bp.pdf.
Palmisano, S., Mursic, R., Kim, J., 2017. Vection and cybersickness generated by head-and-display motion in the Oculus Rift. Displays 46, 1–8. https://doi.org/10.1016/j.displa.2016.11.001.
Peterson, S.M., Rios, E., Ferris, D.P., 2018. Transient visual perturbations boost short-term balance learning in virtual reality by modulating electrocortical activity. J. Neurophysiol. https://doi.org/10.1152/jn.00292.2018.
Porcino, T.M., Clua, E., Trevisan, D., Vasconcelos, C.N., Valente, L., 2017, April. Minimizing cyber sickness in head mounted display systems: design guidelines and applications. In: Serious Games and Applications for Health (SeGAH), 2017 IEEE 5th International Conference on. IEEE, pp. 1–6. https://doi.org/10.1109/SeGAH.2017.7939283.
Rebenitsch, L., Owen, C., 2016. Review on cybersickness in applications and visual displays. Virtual Real. 20 (2), 101–125. https://doi.org/10.1007/s10055-016-0285-9.
Reinhard, R., Rutrecht, H.M., Hengstenberg, P., Tutulmaz, E., Geissler, B., Hecht, H., Muttray, A., 2017. The best way to assess visually induced motion sickness in a fixed-base driving simulator. Transp. Res. F Traffic Psychol. Behav. 48, 74–88.
Schreiber, J.B., Nora, A., Stage, F.K., Barlow, E.A., King, J., 2006. Reporting structural equation modeling and confirmatory factor analysis results: a review. J. Educ. Res. 99 (6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338.
Serge, S.R., Moss, J.D., 2015, September. Simulator sickness and the Oculus Rift: a first look. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting 59. SAGE Publications, Sage CA: Los Angeles, CA, pp. 761–765. https://doi.org/10.1177/1541931215591236. No. 1.
Slipgate Studios, 2017. March 21. Final Rest (VR Experience). [Store page] Retrieved from. https://store.steampowered.com/app/591660/Final_Rest/.
So, R.H., 1999, August. The search forfor a cybersickness dose value. In: HCI (1), pp. 152–156.
So, R.H., Lo, W.T., 1999. March). Cybersickness: an experimental study to isolate the effects of rotational scene oscillations. IEEE In: Virtual Reality, 1999. Proceedings. IEEE, pp. 237–241. https://doi.org/10.1109/VR.1999.756957.
So, R.H., Ho, A., Lo, W.T., 2001b. A metric to quantify virtual scene movement for the study of cybersickness: definition, implementation, and verification. Presence Teleoperators Virtual Environ. 10 (2), 193–215. https://doi.org/10.1162/105474601750216803.
So, R.H., Lo, W.T., Ho, A.T., 2001a. Effects of navigation speed on motion sickness caused by an immersive virtual environment. Hum. Factors 43 (3), 452–461. https://doi.org/10.1518/001872001775898223.
Speicher, M., Feit, A.M., Ziegler, P., Krüger, A., 2018. Selection-based text entry in virtual reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, https://doi.org/10.1145/3173574.3174221.
Stanney, K.M., Kingdon, K.S., Kennedy, R.S., 2002, September. Dropouts and aftereffects: examining general accessibility to virtual environment technology. No. 26 In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting 46. SAGE Publications, Sage CA: Los Angeles, CA, pp. 2114–2118. https://doi.org/10.1177/154193120204602603.
Stone III, W.B., 2017. Psychometric Evaluation of the Simulator Sickness Questionnaire as a Measure of Cybersickness. Doctoral dissertation. Iowa State University.
Takada, H., Fujikake, K., Miyao, M., Matsuura, Y., 2007. Indices to detect visually induced motion sickness using stabilometry. Proc. VIMS 178–183.
Tanaka, N., Takagi, H., 2004. Virtual reality environment design of managing both presence and virtual reality sickness. J. Physiol. Anthropol. Appl. Hum. Sci. 23 (6), 313–317. https://doi.org/10.2114/jpa.23.313.
von Oertzen, T., Brandmaier, A.M., Tsang, S., 2015. Structural equation modeling with Ωnyx. Struct. Equ. Model.: Multidiscip. J. 22 (1), 148–161. https://doi.org/10.1080/10705511.2014.935842.
Watanabe, H., Ujike, H., 2008, December. The activity of ISO/study group on "Image Safety" and three biological effect. In: Universal Communication, 2008. ISUC'08. Second International Symposium on. IEEE, pp. 210–214. https://doi.org/10.1109/ISUC.2008.11.
Wu, T.L., Gomes, A., Fernandes, K., Wang, D., 2018. The effect of head tracking on the degree of presence in virtual reality. Int. J. Hum. Comput. Interact. 1–9.
Young, S.D., Adelstein, B.D., Ellis, S.R., 2007. Demand characteristics in assessing motion sickness in a virtual environment: or does taking a motion sickness questionnaire make you sick? IEEE Trans. Vis. Comput. Graph. 13, 422–428 196, 202, 433.