IOWA STATE UNIVERSITY
Digital Repository

2017

Psychometric evaluation of the Simulator Sickness Questionnaire as a measure of cybersickness

William B. Stone Iii
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd

Part of the Quantitative Psychology Commons, and the Social Psychology Commons

Psychometric evaluation of the Simulator Sickness Questionnaire as a measure of cybersickness

by

William Bruce Stone III

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Human Computer Interaction

Program of Study Committee:
Douglas Gentile, Major Professor
Amy Froelich
Marcus Credé
Jon Kelly
Zlatan Krizan

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

ABSTRACT

Some users of virtual reality (VR) technology experience negative symptoms, known as cybersickness, sometimes severe enough to cause discontinuation of VR use. Despite decades of research, there has been relatively little progress understanding the underlying causal mechanisms of cybersickness. Review of the measures used to assess cybersickness symptoms, particularly the subjective psychological components of cybersickness, indicated that extant questionnaires may exhibit psychometric problems that could affect interpretation of results. In the present study, new data were collected (N = 202) to evaluate the psychometric properties of the Simulator Sickness Questionnaire (SSQ), the most commonly reported measure of cybersickness symptoms, in the context of virtual reality. Findings suggest that the SSQ, as commonly used, is not applicable to VR. An alternative approach to measure cybersickness is suggested. Overall, incidence and severity of cybersickness was very low and participants rated the VR experience as highly entertaining.

CHAPTER 1. INTRODUCTION

The promise of virtual reality is that we can use technology to create immersive experiences of our choosing, independent of our real environment or physical limitations. Without physical constraints, the limits of virtual experience are bound in principle only by imagination. In practice, virtual experiences can have a surprising physical limitation. Just as certain physical motions can cause motion sickness, such as being on a ship in rough seas, sometimes seeing visual motion can lead to feeling unwell, even in the absence of body movement. As virtual reality technology has developed over the past 20 years, some users have reported feeling negative symptoms, similar to motion sickness, after exposure to virtual visual motion. In the context of virtual reality, these negative feelings are known as "cybersickness" (McCauley & Sharkey, 1992).

Aviation and space travel motivated a substantial research history of the causes of motion sickness (Reason & Brand, 1975). As "simulator" and virtual reality technologies were developed in the 1990's, motion sickness research was drawn upon to provide insight to the causes of simulator sickness and cybersickness. Unfortunately, exact mechanisms or mitigations for simulator and cybersickness remain largely unknown (Rebenitsch & Owen, 2016). The slow rate of improvement in understanding cybersickness could be due in part to challenges measuring the psychological construct of cybersickness (i.e., the subjective negative psychological effects of virtual reality exposure). The Simulator Sickness Questionnaire (SSQ) (Kennedy, Lane, Berbaum, & Lilienthal, 1993) is currently the standard measure of cybersickness symptoms. However, aspects of the measurement development process of the SSQ raise doubts

as to whether it should be considered a valid measure of cybersickness. Without a strong measure, it is difficult to conceive of how empirical science can contribute to resolving cybersickness.

Therefore, in order to advance scientific understanding of the mechanisms of cybersickness, further psychometric evaluation of the SSQ as a measure of cybersickness is needed. The current study will collect new data to evaluate the psychometric properties of the SSQ in the context of virtual reality to see if it is an appropriate measure of cybersickness.

1.1. Virtual Reality

The goal of virtual reality (VR) is to immerse the senses with stimuli generated to create the illusion of presence in a simulated (virtual) environment. The terms "immersion" and "presence" are sometimes used interchangeably, however, clarity can be maintained by defining a distinction (Slater & Wilbur, 1997). The term "immersion" will be used to refer to the degree to which interface components generate stimuli that are more or less encompassing of sensory organs. Immersion is an objective quality of a virtual reality system; a video display with a large field of view is more immersive than a display with a restrictive field of view. "Presence" will refer to the degree of perceiving to "be within" the virtual environment; presence is a subjective feeling experienced by the virtual reality user. Generally, immersion can be considered to facilitate presence (Witmer & Singer, 1998).

A "virtual environment" is a computation of a simulated environment translated to generate stimuli that simulate interaction with that environment. Given user inputs, the virtual reality system computes the according stimuli to be delivered from the physical

interfaces of the VR system to the immersed sensory organs. The perception of interaction with a virtual environment is an illusion facilitated by the VR system continually observing user actions, calculating the coinciding virtual environment response, and communicating this information back to the user. From the perspective of the current study, to be a considered a virtual reality system the apparatus must include a mechanism of "natural" interactivity. At the risk of introducing terminology too early in this discussion, "natural" interactivity will be defined as a mechanism of visual motion control that results in concordant visual-motion and vestibular information. This is a distinguishing characteristic of a virtual reality experience and differentiates virtual reality from other immersive experiences, such as watching a stereoscopic movie filmed from the "viewer" perspective, or playing a computer video game in front of a large screen.

Technology

The following description of virtual reality technology will focus on features that characterize current virtual reality technologies, particularly systems recently released for the consumer market. For an overview of the state of the art of virtual reality technologies, see Anthes, García Hernandez, Wiedemann, and Kranzlmüller (2016). The scope of this study of the psychometric qualities of the Simulator Sickness Questionnaire is restricted to virtual reality, and head-mounted display (HMD) virtual reality systems in particular. Given the variety of VR interface and environment combinations, this is a reasonable restriction. HMD VR systems are currently the most accessible, in terms of both cost and availability, as major entertainment companies have recently released HMD VR systems for the mass market. Focusing the current

study on HMD VR systems increases relevance to what will soon become the most prevalent virtual reality experience.

Interfaces (hardware)

Virtual reality interfaces are the physical components of the virtual reality system that the user interacts with. Interface components serve two primary purposes: communication of virtual environment information and capture of user motion, thus facilitating "interaction" with the virtual environment. Some interface components serve both roles, particularly in more sophisticated VR systems.

Display

The display, the interface component that communicates visual information of the virtual environment to the user, is the most defining feature of current VR systems. Non-consumer VR applications in industry and academia have focused on large screen systems, such as Cave automatic virtual environments (CAVEs), where the virtual environment is projected on large, room-scale, surfaces (Muhanna, 2015). CAVEs typically require substantial physical space, display technology, and computing resources. In contrast, consumer VR systems currently use head-mounted displays (HMDs) similar in size and format to ski goggles and can function with consumer-grade desktop PCs (Figure 1, "VRHeads: Field of view face-off: Rift vs Vive vs Gear VR vs PSVR," 2016).

There are several features of HMD design that affect the quality of immersion. In order to create depth cues from binocular disparity, stereoscopic images must be displayed. This requires displaying different images of the virtual scene to each eye. Because the screen is placed a few centimeters from the eyes, lenses are used to

create a more natural focus depth. The lenses can be within one centimeter from the eyes. Current HMDs use one lens for each eye (binocular display) set to a fixed focal distance. To minimize distortion, it is important for each lens to be centered in front of each eye. Given the variation in interpupillary distance, HMDs allow for adjustment of the distance between lenses. In order to save weight, current HMDs employ Fresnel lenses, a type of lens construction that reduces lens thickness by forming concentric rings into the lens material. Unfortunately, many users report that this lens construction results in glare when viewing high contrast images. The concentric rings may also be somewhat visible, distracting from and potentially distorting the display image.

Given the close distance of the eyes to the display screen, a high pixel density display is necessary. Typical HMDs that are currently available employ two display screens, one for each eye, each at a resolution of 1,080 by 1,200 pixels. The display must also refresh at a high rate, otherwise fast head movements would result in a blur of the display image. The current standard refresh rate is 90Hz. However, motion blur can still occur if the display does not utilize a pixel design with low-persistence, that is, individual pixels must be able to change quickly. To meet these requirements, current HMDs utilize organic light emitting diode (OLED) screens.

The last key feature of HMD display design is field of view (FOV). When using a HMD, there are two different fields of view that work together to create the perceived FOV: the display field of view (DFOV) and the virtual field of view (VFOV). Note that the virtual FOV is sometimes also called the "geometric" FOV (Moss & Muth, 2011). DFOV is a function of the HMD design, based on the size of the display screens, lenses, and the distance of the lenses and screens from the eyes. Current HMDs have a DFOV of

around 100 degrees horizontal and 110 degrees vertical ("VRHeads: Field of view face-off: Rift vs Vive vs Gear VR vs PSVR," 2016). For comparison, human field of view is around 180 degrees horizontal and 135 degrees vertical.

Current HMD VR systems utilize the display as the primary stimulus delivery device; however, sound is also an important component of the VR experience. At this time, sound is delivered through standard headphones that connect to the HMD. Some handheld controllers designed for VR also provide generic vibro-tactile feedback. There are a variety of devices under development to deliver more sophisticated sensation through vibration, sound, smell, and even wind (Anthes et al., 2016).

Controls

Control devices facilitate interaction with the virtual environment through capture of user inputs. For current HMD VR systems, the key control device that qualifies the system as "virtual reality" is the HMD itself. By observing user head position, HMD VR systems update the representation of "point of view" in the virtual environment. The position of the HMD in space and angular rotation are captured. Current systems use "outside-in" tracking, where fixed reference points for position are created by additional tracking equipment, such as laser emitters. This requires predefining a physical space for using the HMD VR system and setting up the external tracking equipment. Accelerometers within the HMD supplement the positional tracking with angular rotation information. When combined with low-persistence display screens and a high refresh rate, the experience of head motion to control point of view is very fluid and realistic in current HMD VR systems. Current systems require external equipment to achieve the speed and precision necessary for responsive and convincing interaction, but future

HMD VR systems will likely use "inside-out" tracking, where advanced accelerometers and cameras that are part of the HMD will be able to track the position without the need for additional equipment.

Other interactions with the virtual environment are input through hand controls. Current technology has supplanted joystick and button gamepads, as used for console video gaming, with VR specific hand controller devices that are motion tracked just as the HMD. The design of these controllers varies between system, but the basic form is to use a separate controller for each hand that is somewhat "wand"-shaped (Figure 2, "HTC Vive Pre Installation Guide," n.d.). Buttons are placed along the top, sides, and bottom to be manipulated primarily by the forefinger and thumb. Because the position and rotation of each controller is tracked by the system, the controllers can make use of fewer buttons than traditional gamepads by responding to gestures, that is, predefined movements, and changing button function based on positional context. There are a variety of forthcoming control devices that allow for more natural interaction, such as glove-based handheld controllers, treadmill platforms that detect walking motion, and HMD systems that incorporate eye-tracking (Anthes et al., 2016).

Virtual environments (software)

Virtual environments for current HMD VR systems are based on the development platforms for three-dimensional video games, with modifications to account for the use of the HMD and tracked controllers. Three-dimensional video game development "engines" have established the virtual constructs of mass, dimension, gravity, and point of view. However, there are additional considerations for rendering the virtual environment for stereoscopic display, particularly for HMD use. Just as the HMD lens

position must be adjustable for different users with different inter-pupillary distance, the virtual environment also needs to respond to this change, modifying the distance between the virtual "cameras" accordingly in order to maintain a convincing stereoscopic image. Similarly, the magnification factor of the virtual cameras determines the geometric field of view (GFOV) and should be responsive to the expectations of the user.

Virtual environments for HMD VR systems that rely on external tracking equipment must also accommodate the defined "play area" wherein the VR system components are tracked. Whereas desktop PC and console video games may have had to account for variable display screen and television dimensions, virtual environments must account for the available physical space for the user to move within. Current HMD VR systems require the user to define their physical play area as part of system set up. These physical bounds must be taken into consideration by the virtual environment in order to keep users from trying to interact with virtual objects that are outside the bounds of the physical play area. Currently, virtual environments that require more than one square meter of area are designated as "room-scale". Of course, not all virtual environments can accommodate all physical play area dimensions, and vice versa. Future "inside-out" tracking systems will allow greater flexibility to use VR in arbitrary locations, but could raise new challenges virtual environment developers that may need to accommodate a variety of physical environment dimensions and terrains.

Control of visual motion

For a HMD VR system, the primary immersion vector is visual information displayed through the HMD. Interaction with the virtual environment is primarily

communicated through changes in visual motion depicted by the display. Therefore, in these systems interactivity is fundamentally a function of visual motion control, that is, the correlation between user physical motion and the according visual motion information displayed. In a forthcoming section, possible mechanisms of cybersickness will be discussed. Theoretically, cybersickness may result from conflicting visual and vestibular information. The degree of conflict between visual and vestibular information while immersed in VR depends on the visual motion control mechanics of the virtual reality system and the design of the virtual environment.

In order to increase immersion, presence, verisimilitude, and potentially decrease cybersickness, mechanisms of visual motion control in virtual reality should result in concordant visual and vestibular information. If the user wishes their virtual avatar to look around (changing virtual point of view), they should be able to do so by moving their head. If they wish their avatar to move forward, they should be able to walk forward. Concordantly accommodating virtual navigation beyond the physical bounds of the motion-tracked area requires careful consideration. At this time, the best solution is "teleport navigation", where the user can indicate a destination in the virtual space and instantly transport there. These visual motion control mechanisms should, in principle, coincide with concordant visual and vestibular sensation.

Other control mechanisms of point of view or navigation visual motion, such as using a gamepad to look or move, will cause discordant visual and vestibular stimulation. In lieu of incorporating a motion-base simulation device, any visual motion indicative of indirect locomotion (e.g., driving a car, floating down a river, etc.) will also coincide with discordant visual and vestibular stimulation.

Consider the following virtual reality experiences as characterized by visual motion control mechanism and ordered by degree of visual and vestibular information concordance:

- Completely concordant: HMD, complete position and rotation tracking, navigation through body motion (e.g., "room-scale" HTC Vive, *Waltz of the Wizard*)

- Partial concordance: HMD, complete rotation tracking, but limited position tracking for complete point of view control, navigation control through gamepad (e.g., HTC Vive, *VR Karts)*.

- Minimal concordance: HMD, only rotation tracking for partial point of view control, navigation and partial point of view control through gamepad (e.g., Oculus Rift DK1).

- Completely discordant: HMD, no tracking, point of view and navigation control through gamepad.

From the "information" perspective, any VR experience without completely concordant visuo-vestibular information is "discordant". According to sensory conflict theory, concordant visual motion experiences should not cause visuo-vestibular conflict while discordant experiences should. Therefore, cybersickness should be predicted to occur only after exposure to discordant visual motion control experiences.

It is notable that available metrics indicate that the currently most popular games available on Steam (a popular online video game distribution system) that are exclusively designed for virtual reality utilize only concordant visual motion control (as of September 2016). Estimating popularity of games through available Steam metrics is challenging; however, games such as *The Lab*, *Tilt Brush*, *Audioshield*, *Raw Data*, *The*

Brookhaven Experiment, *Job Simulator*, and *Space Pirate Trainer* are the most highly played and some of the most highly reviewed virtual reality games, and none incorporate discordant visual motion.

The HTC Vive is a new to market, consumer-oriented, HMD VR system co-developed by HTC, a consumer electronics manufacturer, and Valve Corporation, a PC video game developer and software distributor. Valve also owns and develops the Steam video game distribution system. Given their level of involvement and commitment to virtual reality, it is also telling that their showcase VR game, *The Lab*, includes no instances of discordant visual motion. *The Lab* is a collection of minigames of great variety (e.g., archery castle defense, arcade style space shooter, robot repair mechanic), which indicates that the absence of discordant visual motion was purposeful. The complete rationale behind the exclusion of discordant visual motion in the design of these games is unknown, but judging by the extent of user testing typically employed by game developers, it is likely that they found discordant experiences undesirable.

Natural interactivity

The emphasis of natural interactivity as a defining characteristic of a virtual reality experience deserves further explanation as it has been a somewhat underappreciated distinction within the literature and will be used as a criterion for evaluating the applicability of various psychological measures to the context of virtual reality. Interactivity is a distinguishing characteristic of a virtual reality experience because it is a distinguishing characteristic of real experience. From the perspective of human sensation and perception, particularly vision, it is difficult to imagine a real experience

that is truly static. Aside from certain medical circumstances (e.g., temporary paralysis), we always have some degree of interactive control over what we are sensing, and we seem to constantly exert this control, even in minute ways, to interact with our environment. Even when observing a modern human in a common near-catatonic state (i.e., watching television), there are constant small changes in gaze through eye and head movement that correspond with visual motion, that is, changes in our visual field that imply movement. We use visual motion to explore our environment. Changes in visual information from even subtle changes of eye position enable us to distinguish objects by differentiation and identify features such as depth and scale.

Defining "virtual reality" experiences as those that incorporate a "natural" interaction mechanism, that is, a visual motion control mechanism that results in concordant visual and vestibular sensation, is perhaps a luxury of recent technological improvements. However, other researchers have considered head tracking as a criterion for virtual reality (Rebenitsch & Owen, 2014). We visually interact with the real environment by changing the position of our eyes in space. Motion tracking technology has made the natural mechanism of "point of view" control available to all commercial virtual reality systems. Having reached this technological milestone, it is reasonable to define a virtual reality experience as including at least this one mechanism for visuo-vestibular concordant visual motion control. In general, a convincing virtual reality system should respond to the changes we initiate in the ways that we are accustomed to from our real experience, or, in the case of the "fantastical", in the ways that we expect.

Similar, but distinct, contexts to virtual reality

Much of the extant virtual reality literature, particularly that which is focused on cybersickness, has obfuscated some of these distinctions that define the minimum requirements for "virtual reality". Principally, these are not trivial distinctions, and it can make it difficult to discern what research results were relevant to virtual reality or limited to a related context.

The extent of this confusion appears significant. For example, in a recent review of cybersickness research by Rebenitsch and Owen (2016), 65% of the citations purported to be investigations of virtual reality were *not* conducted in the context of virtual reality. The stimulus apparatus used by these studies lacked a mechanism of visuo-vestibular concordant visual motion control. Therefore, these results are not directly applicable to cybersickness as the research context was not virtual reality. This is a theoretically important distinction because of how different visual motion control mechanisms interact with visuo-vestibular information. To further illustrate the differences between virtual reality and other contexts, four distinct visuo-vestibular contexts will be defined: virtual reality, visual motion, motion, and simulation.

Virtual reality

For the sake of clarity and comparison, the context of virtual reality will be defined again. In virtual reality, there is no external force that influences vestibular stimulation; all body motion is self-initiated. Furthermore, there should be at least one mechanism for visual motion control that results in concordant ("natural") visual and vestibular information. There can be visual motion information that imply motion that do not accord to vestibular information (such as a virtual environment of riding a roller

coaster). However, to be considered virtual reality, there should always be some concordant visual motion control mechanism (such as head movement to change point of view in the virtual environment while riding a roller coaster). Degree of immersion is not a defining factor of virtual reality, however, there is intuitively a minimum of visual and head motion immersion necessary to provide concordant point of view visual motion control. Also note that the source of the visual information is not specified; the images displayed could be based on a computer-generated virtual environment, or a camera feed. As long as the displayed visual motion responds concordantly, the visual and vestibular considerations should be the same. For example, the trivial virtual reality apparatus would be a head mounted display with an attached camera feeding the display.

Visual motion

Similarly, in a *visual motion* context, there is no external influence of vestibular stimulation. However, *visual motion* does not require a mechanism for concordant visual motion control. For visual motion, the apparatus and visual information could be identical to virtual reality, but without a mechanism of concordant visual motion control. For example, if a virtual environment of riding a roller coaster experienced through a HMD did not respond to head motion by changing point of view, or a virtual environment navigated by joystick instead of walking around the physical room. The degree of immersion is not a factor in specifying the context of visual motion; displays ranging from HMDs to projector screens to desktop computer monitors could all be included. Computer and console video games would be considered visual motion contexts, as well as fixed-base (non-motion) driving and flight simulators. The optokinetic drum is

also an apparatus for creating a visual motion context (Muth, Stern, Thayer, & Koch, 1996).

Motion

In the context of *motion*, an external force results in vestibular information that may or may not concord to visual motion information. In general, *motion* contexts cause body movement; examples include modes of transportation and travel, such as cars, boats, and planes, as well as centrifuges or other similar apparati. Visual motion and vestibular information could be concordant, such as when driving a car, or discordant, such as when below deck on a boat at sea.

Simulation

In the context of *simulation*, visual motion information is accompanied by concording body motion, resulting in concordant vestibular information. Motion-base driving and flight simulators, where the apparatus ("base") moves the user, would fall into this context. In some ways, *simulation* combines visual motion and motion contexts. Conceptually, the ideal virtual reality apparatus would incorporate both motion-base and natural mechanism for visual motion control.

1.2. Cybersickness

Cybersickness is sickness or general feelings of malaise experienced due to exposure to virtual reality. The term is attributed to McCauley and Sharkey (1992); however, their initial article somewhat confuses cybersickness with *vection*, the illusion of self-motion given visual motion information, a likely related, but distinct phenomenon (Keshavarz, Riecke, Hettinger, & Campos, 2015). In broad terms, cybersickness

symptoms, that is, subjective negative responses, are related to feelings of dizziness, disorientation, eyestrain, fatigue, and nausea. Signs of cybersickness, that is, objective physiological indicators, ostensibly include biometric signs of sympathetic nervous system activity, such as electrodermal activity and heart rate. Emesis (vomiting) would also be considered a sign of extreme cybersickness.

Further terminology should be clarified. The terms "cybersickness", "visually induced motion sickness", "motion sickness", and "simulator sickness" are sometimes used interchangeably. However, each of these terms align with a distinct motion context, distinguished by different combinations of visual motion and vestibular information, and using these different "sickness" terms interchangeably can cause confusion. For example, after observing that passengers of vehicles are more likely to experience motion sickness than drivers, Chen, Dong, Hagstrom, and Stoffregen (2011) investigated whether observers of video game players would experience more "motion sickness" than video game players; their publication uses the term "motion sickness" instead of "visually-induced motion sickness" and makes no reference to the distinction. Similarly, Keshavarz and Hecht (2011) sought to develop a measure of "motion sickness" but tested the measure by exposing participants to a video projected on a screen (a visual motion stimulus). van Emmerik, de Vries, and Bos (2011) wanted to investigate how field of view affects cybersickness, however, they also used a video projected on a screen. Similarly, Liu (2009) investigated "cybersickness" using a desktop video game played on a 19" LCD monitor display.

The present study will not presume that syndromes resulting from exposure to virtual reality, visual motion, simulator, and motion contexts to be interchangeable.

"Cybersickness" will refer to sickness after exposure to virtual reality; "visually-induced motion sickness" (VIMS) will refer to sickness experienced due to *visual motion* contexts; "motion sickness" will refer to sickness due to *motion* contexts, and "simulator sickness" will refer to sickness due to *simulator* contexts.

Due to general symptomatic similarity with motion sickness, a tendency for VR contexts with greater visual motion to result in greater severity of symptoms, and some degree of historical coincidence, investigation of cybersickness has focused on the visual motion aspects of VR and drawn largely on visually-induced motion sickness (VIMS), simulator, and motion sickness research. However, not all potentially aggravating aspects of the VR experience result from visual imagery. For example, the brightness of an HMD screen placed close to the eyes could cause eyestrain, and the restricted field of view and unbalanced weight could lead to strenuous head movement.

There is not a specific theory of the causes of cybersickness. Instead, because of the prominence of the visual motion aspects of virtual reality, reference is made to theories of motion sickness. Motion sickness can result from sufficient externally influenced vestibular sensation, and it seems that the only persons immune are those without a functioning labyrinthine system (Oman, 1990). However, it was noted early on that effect could be mitigated or exaggerated based on the expectation of experience and seemed to be particularly sensitive when the motion context resulted in conflicting visual and vestibular cues. The "sensory conflict theory" of motion sickness posits that it is the discordance of visual and vestibular information that causes sickness. In the context of visual motion, it is the presence of visual motion information and the absence of concording vestibular information that are thought to cause sickness.

Sensory conflict seems to be the most commonly referenced theory in VIMS and cybersickness research (Rebenitsch & Owen, 2016). However, according to Stoffregen and Riccio (1991), sensory conflict cannot explain what contexts are and are not sickness inducing. Instead, they proposed the "postural stability theory" of motion sickness, where sickness is the result of struggling to maintain postural control. Although postural instability has been used as an indicator of cybersickness (Stanney & Kennedy, 1998), few researchers have seemed to consider it the cause of cybersickness. Note, at least one study testing hypotheses of these competing theories failed to find support for the predictions of postural stability theory in a visual motion context (Warwick-Evans, Symons, Fitch, & Burrows, 1998).

Need for a virtual reality specific review

A number of reviews of "cybersickness" are available (Barrett, 2004; Davis, Nesbitt, & Nalivaiko, 2014; Kiryu, Uchiyama, Jimbo, & Iijima, 2007; LaViola, 2000; McCauley & Sharkey, 1992; Nichols & Patel, 2002; Rebenitsch & Owen, 2016; Stanney, Mourant, & Kennedy, 1998; Strauss, 1995; Wilson, 1997). However, all have combined results from research in virtual reality, visual motion, simulator, and motion contexts in a way that makes it difficult to discern what symptoms and effects align to what context. The important differences between these contexts have not been reflected in these reviews of cybersickness. For example, Rebenitsch and Owen (2016) cite an online medical reference guide about motion sickness from the University of Maryland Medical Center as a reference for cybersickness symptoms. Similarly, the description of cybersickness symptoms by Davis et al. (2014) cites LaViola (2000), who in turn cites Kennedy and Fowlkes (1992): *Simulator Sickness Is Polygenic and polysymptomatic:*

Implications for Research, an article about symptoms resulting from exposure to flight simulators, not virtual reality.

Some of this intermixing of research contexts has likely resulted from trying to address rapidly changing technology. Twenty years ago, the distinctions between virtual reality and visual motion contexts were likely too dependent on unique and specialized hardware and virtual environment availability. However, given current availability of high quality virtual reality systems and virtual environments, an updated review of the literature that encompasses only *virtual reality* contexts is warranted. As useful as extant "cybersickness" reviews are, basic concepts of *cybersickness* (in the context of virtual reality) cannot be discerned. A new review of cybersickness literature explicitly scoped to research where the visual motion context was virtual reality, as defined as including at least one visuo-vestibular concordant mechanism of visual motion control, is needed to identify cybersickness prevalence, severity, duration, and even symptoms.

As seen previously, the description of symptoms often provided in cybersickness research literature can be traced back to simulator or motion sickness research. Possibly as a result of historical coincidence, cybersickness is often described by the symptom items that comprise the Simulator Sickness Questionnaire (SSQ), even though there was evidence very early on that participant response to virtual reality differed greatly from simulators (Stanney, Kennedy, & Drexler, 1997). Stanney et al. (1997) observed a very different profile of SSQ responses after exposure to virtual reality compared to the flight simulators the SSQ was designed and validated for. Unfortunately, this did not motivate further investigation at that time into the creation of a

new measure of cybersickness or categorization of virtual reality as a distinct context from simulation. Instead, the SSQ became the standard measure for all manner of contexts that could be associated with motion sickness-like symptoms (Davis et al., 2014; Rebenitsch & Owen, 2016). Consequently, basic understanding of cybersickness symptoms seems to have become conflated with the way it was measured. A thorough, systematic, review of the extant cybersickness literature is warranted.

1.3. Measures of Cybersickness Symptoms

Without a systematic review of the cybersickness literature, the following discussion of measures of cybersickness symptoms should not be considered exhaustive. However, even informal review of extant cybersickness literature indicates that the SSQ is the primary measure of cybersickness symptoms. In practice, the alternative to administering the SSQ appears to be to use a single-item nausea or motion sickness question. As previously mentioned, the SSQ was developed for the context of simulation, not virtual reality. There is one measure that has been developed specifically for the measure of cybersickness, the Virtual Reality Symptom Questionnaire (VRSQ) (Ames, Wolffsohn, & McBrien, 2005). However, it does not appear to have been used in any published cybersickness research. Nonetheless, as the only measure specifically developed for cybersickness, it deserves discussion.

Virtual Reality Symptom Questionnaire

The Virtual Reality Symptom Questionnaire (VRSQ) was developed specifically for the measure of cybersickness (Ames et al., 2005), and appears to currently be the only measure developed specifically with the intention of assessing virtual reality effects. The VRSQ emphasizes the ocular symptoms related to VR exposure, such as

eyestrain. Ames et al. (2005) felt that the SSQ did not adequately address ocular symptoms for application to virtual reality. Unfortunately, there is a significant caveat: the VRSQ has not been validated in the context of virtual reality. The stimulus used in the development study by Ames et al. was a video viewed on a head-mounted display. Given that the video did not allow for any interaction through user input, let alone a mechanism for natural interaction, the study would be considered to have used a visual motion stimulus, not virtual reality. Furthermore, because of scant validity evidence (the development study also tested a very small sample), the measure has seen little use (Davis et al., 2014).

Ames et al. (2005) collected 47 symptom items that had been used in virtual reality research. Items were then reviewed further and those with infrequent participant response from previous research were eliminated. They added an item for vision discomfort to increase the number of items related to vision quality and decided to remove the item "discomfort from eyes" because they felt that it was ambiguous. This left 23 symptom items: 12 nonocular and 11 ocular. They created a seven-option response scale (scored 0-6) with four descriptive labels ("none" (0), "slight" (1, 2), "moderate" (3, 4), and "severe" (5, 6).

In the development study, 16 optometry and vision science students, aged 21 to 28 years donned a HMD (V6 HMD) and viewed a 20-min stereoscopic video (the title of the video was "Eye to Eye"). Due to concern of response bias from pre-exposure administration of the symptom list, half of participants were randomly assigned to complete the VRSQ before exposure, and half did not. Immediately after watching the

video all participants completed the VRSQ and then repeatedly at 2-min intervals for a total of six post-exposure administrations.

No priming effect was found, so they did not differentiate data analysis between experimental priming conditions. Ames et al. (2005) tested item-total correlations at each assessment timepoint to determine whether some items should be discarded. Looking across timepoints, they concluded that only 13 symptom items had a satisfactorily high item-total correlation to remain as part of the measure. Although the nausea item did not meet their criterion for inclusion, they decided to retain it anyway because they felt it may be useful to other researchers using more "dynamic imagery".

The validity of the VRSQ as a measure of cybersickness is questionable given the information provided by Ames et al. (2005). At face value, given that the testing stimulus was not virtual reality, it is reasonable to be reluctant. Their approach to judging the adequacy of the measure lacks psychometric sophistication that should be expected of psychological measures.

A general approach to psychometric evaluation of a multi-item scale, particularly where there is assumed to be more than one latent factor (in this case, the ocular and nonocular item groups), involves conducting an exploratory factor analysis followed by confirmatory factor analysis to determine scale dimensionality and decide what items should be retained. Then, the internal consistency should be analyzed to provide insight into the reliability of the measure. After having established dimensionality and consistency, an evaluation of measurement invariance should be conducted to identify whether repeated administrations of the measure are comparable, that is to say, whether responses collected immediately after exposure can be compared to responses

collected 10 minutes after. Finally, analyses should be conducted to provide evidence for the validity of the measure, such as testing correlation with other accepted measures. The lack of published studies using the VRSQ can likely be attributed to the marginal evaluative information and validity evidence (Davis et al., 2014). Indeed, the most recent review of cybersickness literature by Rebenitsch and Owen (2016) does not reference the VRSQ at all.

Single-item measures

If the SSQ is not reported as the operationalization of cybersickness, it is most common for a single-item scale to be used. The single-item scales share a common approach, asking participants to rate on a zero to ten scale feelings of either nausea, motion sickness, or discomfort, depending on the study. Nalivaiko, Davis, Blackmore, Vakulin, and Nesbitt (2015) used a single-item, 11-option, nausea rating, "0 - no signs" to "10 - ready to vomit". Davis, Nesbitt, and Nalivaiko (2015) used a single-item, 11-option, nausea rating, although the response anchors used were "0 - no nausea/discomfort" to "10 - very nauseous (feeling like vomiting)".

In order to monitor participant safety during exposure to virtual reality, Rebenitsch and Owen (2014) employed an *immersion rating*, where participants were verbally asked *"On a scale of 0-10, 0 being how you felt coming in, 10 is that you want to stop, where you are now?"* (sic) The intention was to discontinue exposure if the participant responded "10". Cybersickness was operationalized by post-exposure SSQ response. Only the point estimate of correlation between SSQ scores and immersion rating was reported; however, based on available information, the 95 percent confidence interval was computed to range between 0.25 and 0.72 (Bonett & Wright,

2000). Fernandes and Feiner (2016) employed this same *immersion rating* question; yet, they decided to interpret responses as a *discomfort score*. It appears that they operationalized cybersickness as both discomfort scores and post-exposure SSQ scores; however, their results report hypotheses in reference to *discomfort* instead of cybersickness. Differences in discomfort scores supported their hypotheses (regarding use of a dynamic field of view change during motion), while SSQ scores did not. They did not report correlations between discomfort and SSQ scores.

1.4. Simulator Sickness Questionnaire

The most common reported measure of cybersickness symptoms is the Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993). The SSQ was developed to measure sickness in the context of simulation and was derived from a measure of motion sickness. To conceptually evaluate the applicability of the SSQ to cybersickness, a number of aspects of the development, design, usage, and prevalence of the SSQ will be discussed.

Precursor measures

Two measures of motion sickness lead directly to the development of the SSQ; describing these measures will establish relevant historical context.

Pensacola Diagnostic Rating Scale (PDRS)

The Pensacola Diagnostic Rating Scale (PDRS) is a measure of motion sickness (specifically for *motion* contexts). The PDRS was used by NASA to assess space sickness (with varying methodology) between 1973 and 1984, at which time a new questionnaire was developed at the Johnson Space Center. The PDRS preceded and

inspired the Motion Sickness Questionnaire (MSQ), which lead to the SSQ. There does not seem to be a specific reference that details the development of the PDRS, but Stout and Cowings (1993) credit Graybiel, Wood, Miller, and Cramer (1968); Kennedy, Tolhurst, and Graybiel (1965); Miller and Graybiel (1970); and Miller and Graybiel (1974). The following description of the scale and usage is based on Stout and Cowings (1993).

The PDRS is a revised symptomatology scale from "Slow Rotation Room" (SRR) motion sickness research. Scale items include: vomiting, increased warmth, dizziness, headache, drowsiness, sweating, pallor, nausea, epigastric discomfort, epigastric awareness (see Table 2 of Stout and Cowings, 1993). Recorded responses combined both participant and experimenter evaluation of participant symptoms, with varying response options. For example, headache was rated by the participant on two levels (present or absent), drowsiness on three levels (mild, moderate, or severe), and nausea by a combination of three items (epigastric awareness, epigastric discomfort, and nausea), each on three levels (mild, moderate, or severe). Pallor was assessed by an experimenter on three levels (unclear rating levels). A total score was computed by weighted sum of item responses based on symptom type and intensity; score ranges were categorized by severity of motion sickness: 1 to 4 mild, 5 to 7 moderate, 8 to 15 severe, and greater than or equal to 16 "frank" motion sickness.

Pensacola Motion Sickness Questionnaire (MSQ)

The SSQ was derived directly from the Pensacola Motion Sickness Questionnaire (MSQ). There does not seem to be a specific reference that details construction and usage of the MSQ, but Kennedy et al. (1993) credit Kellogg, Kennedy,

and Graybiel (1965); and Kennedy et al. (1965). The MSQ consists of a list of 25 to 30 symptoms; Kennedy et al. (1993) list 28 symptoms (see Table 1, Kennedy et al., 1993). Symptom items are rated on four levels, "none" (0), "slight" (1), "moderate" (2), and "severe" (3). Other details of the development of the MSQ are indeterminable from available sources; however, Kennedy et al. (1993) note that the MSQ was developed in context of motion stimuli sufficiently severe to induce emesis or near-emesis. A total score was computed by summing item scores; however, the final scoring was determined to specify emesis as the highest possible score; this is known as a "configural" approach to scoring (Kennedy et al., 1993). That is to say, no matter the other item responses, if a participant vomited they were rated with the highest score. No scale dimensionality, reliability, or validation information is available.

Development of the SSQ

The development of the SSQ was published in *Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness (* Kennedy et al., 1993). The SSQ is a selection of 16-items from the MSQ with a different scoring scheme. The scoring scheme is based on factor analyses resultant from administrations of the MSQ in the stimulus context of simulation. Kennedy et al. (1993) conducted a factor analysis on MSQ data collected from a previous study of simulator sickness experienced by US Navy pilots after exposure to flight simulators (Kennedy, Lilienthal, & Berbaum, 1989). In Kennedy et al., (1989), simulator sickness was operationalized by responses to a 28-item MSQ. In Kennedy et al. (1993), the authors recognized that there could be a different latent variable structure for MSQ responses in the context of a simulator because they found that sickness experienced after exposure

to a simulator rarely resulted in emesis, in contrast to the motion-based contexts that the MSQ was developed for, which often evoked emesis.

Validation study

In the validation study of the SSQ, participants were military pilots, however, the exact number of participants can not be determined (Kennedy et al., 1993). Responses from some pilots were collected multiple times; how many responses were from the same pilots was not known. The 28-item MSQ was administered pre and post approximately 1,200 simulated flight sessions. Ten different flight simulators were used. Details of the simulator designs were published in Kennedy et al. (1989). Five of the simulators seemed to elicit little to no symptomatic response, therefore, the authors decided that only post-exposure responses from the other five simulators that did seem to elicit a "significant" symptomatic response would be used, leaving approximately 600 assessments to be used for the evaluation of dimensionality of what would become the SSQ.

Dimensionality and scoring

Items with low response frequency were eliminated (frequency less than 1%), as well as those that did not seem to coincide with other symptoms, leaving 16 items for analysis. To determine dimensionality, two different factor analyses were conducted. First, a principal-factors analysis was conducted, followed by normalized varimax rotation. A three-factor, partially independent, solution was chosen after comparing three-, four-, five-, and six-factor solutions (no rationale for selection of a three-factor solution was provided).

Kennedy et al. (1993) felt that the varimax-rotated factor structure indicated the presence of a general factor. They then conducted a second factor analysis, a hierarchical factor-analysis, and concluded that there was a general factor. However, because they felt that the varimax-rotated factor structure would be more robust, and because an unknown percentage of responses were not independent (i.e., some responses were from the same participants), they determined that scoring of the SSQ would use the varimax-rotated factor structure, not the hierarchical.

Three sub-factors were identified and named based on the sub-factor item with greatest varimax loading (Figure 3). The sub-factors are referred to as Nausea (*N*; "Nausea" item), Oculomotor (*O*; "Eyestrain" item), and Disorientation (*D*; "Dizzy (eyes open)" item). Scoring of each sub-factor is based on the varimax-rotated factor structure. Items included in the scoring of a sub-factor have a varimax loading greater than .30 for that sub-factor. Consequently, each sub-factor consists of seven items. It should be noted that some items are included in the scoring of more than one sub-factor, e.g., "Difficulty focusing" is included in sub-factors *O* and *D*, "Nausea" (*N* and *D*), "Difficulty concentrating" (*N* and *O*), and "Blurred vision" (*O* and *D*).

The sum of the raw scores for the items included within a sub-factor is then multiplied by a constant, which differs for each sub-factor, to scale the sub-factor score to have a standard deviation of 15 for all 1,200 observations (Figure 4). Sub-factor scores can range from 0 to 30.54 (N), 28.58 (O), and 34.92 (D). A score of total severity (TS) is derived by summing the raw (unscaled) sub-factor scores, then multiplying that sum by 3.74. Total severity scores can range from 0 to 235.62.

Prevalence and usage

A systematic review of the usage of the SSQ in the context of virtual reality is warranted. This review would focus on published research where the SSQ has been used as a primary dependent measure of cybersickness, but would include all studies of virtual reality that have reported using the SSQ. The primary objective of the review would be to establish how often the SSQ is used and how it is used, such as how many times it is administered before, during, and after exposure to VR. Furthermore, information about what scores are reported among the four SSQ factors (total score and three subfactors) and how they are tested could be gathered.

Critiques of the SSQ as a measure of cybersickness

Stanney et al. (1997) indirectly critiqued the SSQ as a measure of cybersickness by noting that responses to the SSQ after exposure to virtual reality demonstrated a distinct pattern from responses collected after exposure to flight simulators. The symptom profile and intensity (as operationalized by SSQ responses) was greater after exposure to VR across a range of experimental contexts and research sites, with higher disorientation responses in particular. Stanney et al. concluded that this differentiation in response profiles implied that cybersickness was a distinct phenomenon from simulator sickness and warranted context specific investigation.

Given the observed variation in SSQ responses between simulator and VR contexts, and the absence of a specific validation study of application of the SSQ to VR, it would seem reasonable to question the validity of the SSQ as a measure of cybersickness. However, this conclusion was not discussed by Stanney et al. The

SSQ would go on to become the de facto measure of cybersickness despite this early evidence that should have raised validity questions.

Ames et al. (2005), a group of optometrists, criticized the SSQ for not having a wide enough range of ocular specific symptom items, particular for assessing effects of HMD usage. They also pointed out that it may be beneficial to include fewer items so that the measure can be completed faster due to the speed that cybersickness symptoms can change (they suggest post-exposure symptoms could decrease to minimal levels within five minutes).

Rebenitsch and Owen (2016) list a few disadvantages of the SSQ that keep some researchers from using the measure. First, because subfactors are non-exclusive, meaning that multiple SSQ items are included in more than one subfactor, interpretation of results can be limited and it can be difficult to compare differences between subfactor scores. Second, they criticize the SSQ for being overly sensitive. In fact, they claim that "merely closing one's eyes for an extended period of time can register on the measurement"; however, this claim is uncited. Finally, the factor structure established from responses collected from a sample of military pilots may not be generalizable to the general population.

Pretesting effect

Young, Adelstein, and Ellis (2006) conducted a study specifically to critique what they hypothesized to be a possible response bias effect through pre-exposure administration of the SSQ. Although Kennedy et al. (1993) specified that only postexposure responses should be scored (see "Using and scoring the SSQ: Administration", p.211) and recommended against testing of difference scores between

pre- and post-exposure SSQ responses (p.206), many studies have ignored this and administered the SSQ both before and after exposure to the stimulus. Even when administered pre-exposure, sometimes difference scores are tested and sometimes only post-exposure scores are used. Young et al. (2006) hypothesized that pre-exposure administration of the SSQ could create response bias in post-exposure SSQ responses, specifically in the form of a "demand characteristic". Young et al. speculated that after completing the SSQ before exposure (a "pretest") a participant could infer that the experimenter expected that the stimulus should cause them to respond differently to the second SSQ than the pretest SSQ. To satisfy the perceived expectations of the experimenter, participants report increased symptom severity in the posttest SSQ, even though they are not feeling any worse.

To investigate the pretesting effect of the SSQ, Young et al. (2006) conducted an experiment and found that post-exposure SSQ responses were 80% higher for a group that completed the SSQ pre-exposure compared to a group that completed the SSQ only post-exposure. Young et al. concluded that the observed difference in posttest SSQ responses was due to demand characteristics; however, the design of their experiment can not eliminate alternative explanations. It is possible that exposure to the pretest SSQ caused a "nocebo" effect, where the pretest SSQ caused participants to expect to become ill during immersion, and such expectation caused symptoms to manifest. As discussed by Young et al., a nocebo effect differs from a placebo effect in that nocebo effects are negative effects (e.g., symptoms of nausea) and placebo effects are positive (e.g., decrease in nausea). Although further investigation of the pretesting

effect of the SSQ is warranted, the results of Young et al. indicate further challenges to the interpretation of SSQ responses.

Revisions and refactorings of the SSQ

To address some of the criticisms of the SSQ, three separate research groups, none directly connected to the original SSQ development group, have published revised versions of the SSQ based on their own data collection and analysis. Note that none of the following revisions of the SSQ have been widely adopted by cybersickness researchers; published SSQ scores are almost always based on the original approach developed by Kennedy et al. (1993).

Revised-SSQ (RSSQ)

Kim, Parker, and Park (2004) developed a revision of the SSQ that they call the Revised-SSQ (RSSQ). They had several motivations for revising the SSQ. They felt that the SSQ was missing some potentially important symptoms, such as facial pallor, ataxia, and vomiting. The scoring scheme weights each symptom equally, which may not align with the theoretical importance of the symptom. Furthermore, they felt that because the scoring did not account for baseline response, post-exposure scores may only be interpretable for healthy individuals, which they claimed the general population is not. They also pointed out that the ordinal response options make interpreting scores somewhat difficult. They were also concerned about the generalizability of the original development study sample of military pilots to the general population.

A list of the original 28 MSQ items and 3 additional items, pallor, difficulty equilibrating, and muscle stiffness from strain, was presented to a panel of 15 simulator sickness experts. The experts reduced the list to 24 items and assigned variable

weights depending on the importance of the symptom. Instead of using the original 4-option ordinal response scale, they decided to use an 11-option 0 ("nothing") to 10 ("very severe") scale. This scale was presented with 11 response options, but with labels only at the two extremes.

The RSSQ was administered to 64 adults (12% women) before and after a fixed-base driving simulator on two separate study sessions, yielding 128 pre- and 128 post-exposure observations. The display screen was a 24 inch desktop PC monitor. Exposure time for each session ranged between 15 and 20 minutes. A factor analysis was conducted on "128 data sets", although it is not clear whether they used only RSSQ responses from one session or either pre- or post-exposure. The details of the factor analysis are vague; however, they reported a varimax rotation was performed. They decided the analysis indicated a four-factor model, although no rationale for this choice was provided. The four factors were labeled "nausea", "disorientation", "ocular discomfort", and "strain/confusion". The scoring takes into account the varimax factor loading, the difference of post-exposure to baseline, and weights according to the expert assessment of the importance of the symptom item. This scoring is not straightforward, but a detailed explanation of the scoring procedure is provided.

The RSSQ has seen limited use (Serge & Moss, 2015). There are a few reasons why the RSSQ has not replaced the SSQ in practice. First, the development study sample, although more representative than Kennedy et al.'s (1993) sample, was very under-representative. Also, the scoring procedure is likely intimidating to many researchers. Psychometrically, there is too little information provided about the measure development process, particularly in regard to factor structure selection and

reliability. It does not appear that the multilevel structure of the responses were accounted for in the analysis. Based on the results of Young et al. (2006), the administration of the RSSQ pre-exposure could have significantly inflated the post-exposure scores, which could have an unknown effect on the resulting factor structure. No general validity information was provided, and there is no available validity information available for the context of virtual reality. Given these concerns, and that the measure was developed in a visual motion context and not virtual reality, there is not sufficient reason to believe that the RSSQ would be any more applicable as a measure of cybersickness than the original SSQ.

Bouchard et al. refactored SSQ

Bouchard, Robillard, and Renaud (2007) conducted a study to create a cybersickness specific refactoring of the SSQ. They were motivated to collect new data and test for an alternative factor structure for reasons similar to previously mentioned critiques. First, the Kennedy et al. (1993) factor structure includes some items on multiple subfactors, therefore biasing the computation of the total score. Second, because the SSQ development study used responses from military pilots after exposure to flight simulators, the factor structure may not be applicable to the general adult population after exposure to virtual reality.

SSQ responses were collected from 371 adults (71% female) before and immediately after exposure to virtual reality. The sample was also distributed between those with a DSM-IV diagnosed anxiety disorder (n = 164) and "normal controls". A variety of virtual reality technologies were used, including HMDs and CAVE-like

systems, and also a variety of virtual environment contexts. Participants were immersed for a range of times, ranging between 5 and 60 minutes.

The results of a principal factor analysis indicated a two-factor model with no items loading on more than one factor. All 16 original items were retained and the subfactors were labeled as *Nausea* and *Oculomotor*. They replicated the factor analysis on the anxious and normal control subgroups and found support for the same two-factor structure.

Bouchard et al.'s (2007) refactoring has not seen much use. The study sample, predominantly female with a much higher incidence of anxiety than the general population, is still unrepresentative. This may be the primary reason this refactoring has not seen much use (Rebenitsch & Owen, 2016). Bouchard et al. also misinterpreted Kennedy et al. (1993) and administered the SSQ pre-exposure. Similar to Kim et al. (2004), this could have biased post-exposure responses. Furthermore, the factor analysis approach did not take into consideration the ordinal structure of the SSQ responses. Although the motivations were in the right place, there is not evidence that Bouchard et al.'s refactoring of the SSQ is any more applicable to the measure of cybersickness than the original factor model.

Bruck and Watters refactored SSQ

Bruck and Watters (2011), "*The factor structure of cybersickness,*" were motivated to address the factor structure of the SSQ in order to try to incorporate physiological response. The SSQ was administered to 28 adults (36% female) before and after exposure to a visual motion context (a video of a roller coaster-like ride was projected on a screen). It is not clear how long participants were exposed to the visual

motion display. Respiration and electrocardiogram (ECG) data were also captured; respiration and heart rate effects were operationalized as the average values from the final minute of exposure.

A principal components analysis, without rotation, was conducted to evaluate factor structure of the responses. This analysis included respiration and heart rate as variables alongside the 16 SSQ item responses. After observing the number of eigenvalues greater than one from the correlation matrix, a four factor structure was chosen. They labeled the four components "cybersickness", "vision", "arousal", and "fatigue".

Bruck and Watters (2011) refactoring has had limited use. The validity of the analysis is questionable for a few reasons. First, the sample size of the study, although more representative of gender than some, was still underpowered given general guidelines for factor analysis (Mundfrom, Shaw, & Ke, 2005). Secondly, they conducted a principal components analysis (PCA) instead of an analysis of the common factor model, such as exploratory or confirmatory factor analysis (EFA or CFA). Although some researchers consider these analyses equivalent, they have very different purposes and cannot be used interchangeably (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Given the intention of identifying latent constructs underlying cybersickness (the "factor structure"), PCA was not the correct analysis approach and likely biased conclusions. There is no evidence that Bruck and Watters' refactoring of the SSQ is any more applicable to the measure of cybersickness than the original factor model.

Psychometric review of the SSQ

Having discussed a selection of published critiques and responses to the SSQ and failing to identify a clear alternative, a more thorough psychometric review is warranted. There appear to be several notable psychometric concerns with the SSQ as a measure of cybersickness. These critiques can be grouped by those relating to methodology, analysis, and validity. On the whole, it is clear that new data need to be collected with the specific intent of evaluating the validity of the SSQ as a measure of cybersickness.

Methodological critiques

In order to validly interpret SSQ scores as a measure of cybersickness, there is need for a more representative validation sample and testing context. The tolerance and responsiveness of military pilots to motion and visual motion sickness, among other characteristics, are likely not representative of a general population. Careful selection of a study sample is critical to draw informative conclusions from factor analysis (Fabrigar et al., 1999). Furthermore, the study stimulus should be virtual reality. As discussed in previous sections, simulation, both fixed- and motion-base, contexts lead to a significantly different combination of visual motion and vestibular information than virtual reality.

The SSQ only includes symptom items. Although the diversity of items appears to allow dimensional assessment of symptoms, there are other subjective effects of virtual reality experience that are not being measured. Ostensibly, the intention of engaging in virtual reality is to have a positive, if not exciting, experience. By combining negative symptoms and positive items, such as "energetic", not only could a more

complete picture of the subjective effects be captured, but presenting a mix of positive and negative items could decrease the response bias reported by Young et al. (2006) by masking the "intention" of the questionnaire.

The administration of the SSQ repeatedly, typically before and after exposure to the motion-related stimulus, introduces both methodological and analytical problems. As discussed previously, a pre-exposure administration appears to bias post-exposure responses. Perhaps creation of a separate version of the measure specifically for baseline assessment could address response bias (Morean, Corbin, & Treat, 2013). However, a separate baseline measure would not address repeated administrations of the SSQ during exposure (for example, Graeber and Stanney, 2002). The demand characteristic response bias of a pre-exposure administration is likely exacerbated by repeated measures during exposure. A "during exposure" specific version of the measure could be constructed, but it may be more reasonable to consider changing the framing of items to minimize response bias.

Graeber and Stanney (2002) demonstrated another methodologically confounding approach to using the SSQ: verbal administration of the symptom items by the experimenter. Given possible demand characteristic response bias, verbal administration by the experimenter is possibly the most biased administration approach. Every effort should be made to administer the SSQ in a way that encourages the participant to feel that their responses are anonymous.

Analytical critiques

The following analytical critiques apply most directly to the factor analysis conducted by Kennedy et al. (1993); however, some of these concerns also apply to the

revisions and refactorings of the SSQ discussed previously. First, scale dimensionality was developed with only half of the responses (responses from the low-sickness simulators were not used); however, all responses were included to determine subfactor measured variable loadings. This may have biased the results. A more reasonable approach would have been to conduct an exploratory factor analysis (EFA) on responses from the high sickness portion of the sample, then test the resulting model on the low sickness portion of the sample through confirmatory factor analysis (CFA). If the study sample size is large enough, it can be advantageous to randomly divide responses into two groups, then conduct an EFA on one half and test the resulting model on the second half using CFA (Fabrigar et al., 1999). Given the large sample size of Kennedy et al. (1993), it is disappointing that they did not follow this approach.

As mentioned previously, responses to SSQ items are collected using a four-option response scale; however, scales with ordinal items require specific consideration for factor analysis (Baglin, 2014). For example, using a polychoric correlation matrix instead of a Pearson correlation matrix, which would account for ordinal responses that represent normally distributed underlying response variables. There is no indication that the ordinal rating of items was considered in the factor analysis.

An orthogonal rotation (varimax) was used; however, it is likely that the latent constructs of interest are correlated. An oblique rotation method is almost always a better choice for factor analysis of psychological constructs (Fabrigar et al., 1999). Given that the reported factor model includes multiple items with significant loadings on multiple factors, it is likely that an oblique rotation method would have yielded more robust results.

No rationale for choosing a three-factor model was provided. The decision process for determining the number of factors has important implications on the validity of conclusions (Fabrigar et al., 1999). It is impossible to judge the reasonableness of this decision process without exposition.

No rationale for the scoring of total severity was provided. As mentioned by Bouchard et al. (2007), because scores for total severity sum the raw-scores for each sub-factor, total severity is weighted toward items that are included on multiple sub-factors. This scoring approach makes it difficult to interpret reliability as the total score includes some items more than once.

No assessment of reliability was provided. Estimating reliability is important to identifying robust solutions. The low communalities of some items reported by Kennedy et al. (1993) may indicate unreliable measurement or items that are not relevant to the latent constructs of interest.

No measurement invariance evaluations of the SSQ have been reported. Following from classical test theory, the measurement model of a latent construct incorporates a complex set of assumptions that require evaluation beyond typical reliability and validity evidence when comparing responses between groups (Vandenberg & Lance, 2000). When considering applications of the SSQ, there are a number of different comparative groups at different levels of interest. The most obvious comparisons are between responses collected after exposure to different visual and motion stimuli. As previously discussed, exposure to virtual reality, visual motion, simulation, and motion contexts result in significantly different combinations of visuo-vestibular information. The robustness of the SSQ as a measure of feelings of malaise

resulting from exposure to these different contexts cannot be taken for granted and a formal evaluation of measurement equivalence is need. Similarly, it cannot be presumed that the scores collected repeatedly over a study session are equivalent (Morean et al., 2013). Given that symptoms follow a time-dependent onset and decay sequence, it is a non-trivial question as to whether the SSQ is measuring the same construct at each timepoint. There is also indication from extant results of variable susceptibility to cyber-, visual motion-, simulator-, and motion sickness in the general population, with some responding strongly and some not at all. Again, it cannot be assumed that SSQ scores are equivalent between different susceptibility groups.

Validity critiques

Evidence of the validity of the SSQ as a measure of cybersickness is lacking. Experimental studies that have manipulated features of the virtual reality experience to affect cybersickness could be considered to provide criterion-related validity, yet, such results are questionable without first evaluating the measurement equivalence of the SSQ in the context of virtual reality. Even if SSQ scores are assumed to measure cybersickness, criterion-related and discriminant validity evidence appears ambiguous.

Cobb (1999) did not find that SSQ responses correlated with postural stability. There is not evidence that the SSQ correlated well with the *immersion/discomfort* rating used by Rebenitsch and Owen (2014) and Fernandes and Feiner (2016); the computed 95 percent confidence interval covered a wide range of modest correlations, ranging between .25 and .72. Furthermore, Fernandes and Feiner reported that their dynamic depth of field manipulation resulted in lower discomfort rating scores but not SSQ scores.

Limited concurrent validity evidence is available. Plouzeau, Paillot, Chardonnet, and Merienne (2015) found that participants exposed to proprioceptive vibrations during "walking" navigation in the virtual environment, controlled by gamepad, had lower SSQ scores and higher postural stability. Rebenitsch and Owen (2014) reported that SSQ scores can be predicted by video game play and motion sickness history.

1.5. Open Questions in the Study of Cybersickness

Thus far, the exposition of extant research of cybersickness has identified important open questions. First, the description of cybersickness seems confounded with how it has been measured. Instead of the experience informing the measurement, it seems that the measurements have informed the description of the experience. At this time, there is evidence that some users of VR have a negative experience, but empirical description of the characteristics of that experience is lacking. Some definitions of cybersickness that appeal to specific subjective physiological symptoms seem to be empirically unfounded. Operationally, what remains is a vague understanding of cybersickness as "unwanted response to virtual reality exposure." Although this definition is not *trivial*, it certainly is not as informative to investigation of underlying mechanisms as a definition that identifies particular symptom constructs.

Relatedly, there seems to be a dearth of cybersickness research, and studies of visually-induced motion sickness (VIMS) seem to often be confused for cybersickness. As discussed previously, the mechanisms of VIMS, theoretically caused by discordant visual and vestibular information, are principally absent in virtual reality. Without further investigation, it should not be taken for granted that cybersickness is related to conflicting visual motion and vestibular information. Therefore, there is not convincing

extant evidence of the character nor prevalence of cybersickness. Further inquiry of subjective response to virtual reality is needed to create an empirical definition of cybersickness and description of prevalence.

Given recent technological advancements, new empirical inquiry of subjective response to VR is even more important. Although the historical trend of technological advancement has not indicated a corresponding decrease in cybersickness (Rebenitsch & Owen, 2016), recent VR systems such as the HTC Vive are exceedingly more advanced than systems previously available. Due to the recent release of the HTC Vive and similar systems, empirical evaluation of subjective response is not yet available.

Use of the Simulator Sickness Questionnaire in the context of virtual reality appears problematic. Despite popularity, there is insufficient evidence to support the validity of the SSQ as a measure of cybersickness. Furthermore, there is not evidence that any other extant measure of symptoms of cybersickness, including refactorings and revisions of the SSQ, offer an obvious advantage to the SSQ. Therefore, there is currently no empirically valid measure of cybersickness. Without a valid measure, scientific investigation of cybersickness is challenging. An argument could be made that a new measure should be created immediately. However, because the majority of published studies have operationalized cybersickness by responses to the SSQ, further empirical evaluation of the SSQ should be conducted first (absence of evidence is not evidence of absence). If the SSQ is found to be inadequate, a convincing empirical argument will be needed to persuade researchers to abandon the established research paradigm utilizing the SSQ. Therefore, there is a need for evaluation of SSQ responses

collected in the context of virtual reality, both to test the appropriateness of the standard SSQ factor structure and to identify other possible factor structures.

Further evaluation of the SSQ does not, however, necessarily preclude inquiry toward the development of a new measure of cybersickness. Several of the criticisms of the SSQ as a measure of cybersickness can not be resolved by adjustment of factor structure alone. Given that empirical descriptions of the the phenomenon of interest, cybersickness, are absent, best practice for measurement development would suggest to begin with collection of qualitative responses to VR exposure. An informative example of a measurement development process can be found in Morean et al. (2013). Morean et al. developed a measure of the subjective effects of alcohol exposure, Subjective Effects of Alcohol Scale (SEAS). A particularly relatable component of the SEAS is that the measure takes into account both the positive and negative subjective experiences of alcohol exposure. For example, as alcohol consumption increases, individuals may experience a spectrum of reactions, from happiness and relaxation to sadness and aggression. Importantly, these emotions are not necessarily exclusive and can change as alcohol absorption levels change across the exposure period. Including items in the SEAS measure that address both positive and negative responses greatly extends information available in collected responses.

There are further methodological benefits by including both positive and negative items. For example, a measure that includes both positive and negative items is less prone to demand response bias where the respondent infers the response that the experimenter wishes to observe through inadvertent "cues" such as item wording and context. This would directly address the sort of demand characteristics identified from

pretest administration of the SSQ (Young et al., 2006). Following the example of the development of the SEAS, development of a new measure of the Subjective Effects of Virtual Reality (SEVR) would begin with qualitative data collection that would inform the creation of a large set of candidate scale items. The candidate items would cover the range of experience description terms collected from the qualitative assessment.

Comparative judgment of the informative quality of responses to the SSQ and a new SEVR scale would be based on evaluating relations with variables that *should* imply construct and convergent validity. However, given the dearth of cybersickness research, "should" is seen from an exploratory rather than confirmatory perspective. If the mechanisms of cybersickness are similar to VIMS, then motion sickness history should correlate with cybersickness. Yet, because VR is a novel experience to the general population, the idea of experiencing VR may create anxiety of motion-sickness-like experience in those with motion sickness history. Some people have experienced VIMS after exposure to "first-person" video games and may presume that VR would cause a similar response, potentially manifesting a self-fulfilling prophecy of negative VR experience. Assessing and managing such expectations is an important consideration for conducting cybersickness research.

As a psychological construct, measurement of cybersickness entails many challenges. Identifying objective indicators, such as physiological signs, would be very useful to establishing valid operationalization of cybersickness. To the extent that cybersickness is understood to be a "discomforting" subjective response to VR, physiological indicators of a stressful response should coincide with report of cybersickness. If cybersickness is related to visual motion information, then

cybersickness should coincide with physiological indicators of motion sickness, such as electrodermal activity and postural stability (Chen et al., 2011; Dong & Stoffregen, 2010; Hakkinen, Vuori, & Paakka, 2002; Shupak & Gordon, 2006; van Emmerik et al., 2011).

1.6. Objectives of the Present Study

The present study will address many of the aforementioned questions. Taken together, these questions indicate that cybersickness research is primarily in an "exploratory" phase, thus, most of the following objectives are exploratory. The research objectives of the present study are to:

- Psychometrically evaluate the Simulator Sickness Questionnaire in the context of virtual reality. Specifically, it is hypothesized that the Kennedy et al. (1993) factor model will not fit SSQ responses in the VR context,

- Evaluate the potential of alternative measure of cybersickness, potentially based on SSQ items or report of attitudes about the VR experience, and

- Evaluate a new measure of subjective evaluation and attitudes of the VR experience, and explore general subjective responses to state-of-the-art, consumer-oriented, VR systems and the potential of physiological measures as indicators of cybersickness.

Some critiques of previous SSQ evaluation studies will be addressed directly. Participants will be recruited from a more representative study population. Sample size requirements will be determined beforehand. Part of the evaluation process will be to establish concurrent and discriminant validity through correlation of SSQ responses with evaluations of the virtual reality experience and potential physiological indicators of cybersickness. Psychometric evaluation of the resulting responses will follow best

practices. The measurement of VR experience will directly address the limitation of the SSQ only including subjective physiologically-oriented items. In contrast to the SSQ, the experience measure items will cover a diverse array of subjective psychological effects, such as enjoyment and frustration, of which physiological symptoms are only a subset. The physiological indicator of interest is electrodermal activity (EDA). Although EDA has been shown to relate to motion sickness, such a relationship has not been empirically evaluated in the context of virtual reality.

CHAPTER 2. METHOD

To evaluate the psychometric properties of the SSQ in the context of virtual reality, a study was conducted to collect participant responses to the SSQ after exposure to virtual reality. Participants completed other measures that should correlate with SSQ response, such as evaluation of the VR experience, motion sickness history, video game experience, and prior virtual reality experience. Physiological responses were also collected, such as those indicative of stressful sympathetic response (e.g., electrodermal activity and heart rate). SSQ responses were collected immediately after a 20 minute virtual reality session.

As demonstrated by Young et al. (2006), it is important to minimize response bias when administering the SSQ. Therefore, in the present study, the SSQ was not administered prior to virtual reality exposure. Furthermore, to minimize participant expectation, the study was framed to potential and participating participants as an evaluation of video games in general, with no direct mention of investigation of cybersickness or adverse symptoms. For the benefit of external validity, the virtual environments used were commercially available virtual reality games. In order to minimize stimulus sampling limitations, a set of games were chosen with a variety of control mechanics and game contexts (e.g., a "wizard" simulator and a roller coaster simulator).

The primary analysis of the SSQ responses will be a confirmatory factor analysis (CFA) based on the factor structure published by Kennedy et al. (1993). If the initial CFA results are contrary to the Kennedy et al. factor structure, the dataset will be randomly split with the intention of conducting standard exploratory factor analysis

(EFA) on one half, then testing any resulting factor model on the other half through CFA. Therefore, the potential split-half, EFA-CFA, sequence is the primary determiner of the needed sample size.

Determination of adequate sample size for the factor analysis followed the guidelines of Mundfrom et al. (2005). Recall that the SSQ is composed of 16 items and that the Kennedy et al. (1993) factor structure includes three factors with communalities ranging from .16 to .66. Therefore, the items to factors ratio is 5.33 and the communality is considered "wide". Referring to Mundfrom et al. (2005), a sample size of 200 should allow for "excellent" agreement with population solutions for the initial full-sample CFA and "good" agreement for potential subsequent split-half, EFA-CFA analysis.

2.1. Participants and Design

Students enrolled in introductory psychology and communication studies courses at Iowa State University (ISU), a public university of the state of Iowa in the midwest United States, were invited to participate in the study. Of the 30,671 undergraduate students enrolled at ISU in the Fall semester of 2016, 43% were female ("Iowa State Gender, Ethnicity & Residence Reports," n.d.). The majority of students were White (87% White, 3% Asian, 5% Hispanic/Latino of any race, 2.6% Black, 2.3% Two or more races, 0.2% American Indian or Alaskan Native, 0.08% Native Hawaiian or Pacific Islander). The majority of students were residents of the state of Iowa (62%). Inferences made from the data collected from the study sample apply to the study population of introductory psychology students at ISU. Introductory psychology students at ISU consist of students taking Introduction to Psychology, Developmental

Psychology, or Social Psychology. Students in these courses are required to either engage in research as a participant or complete quizzes about published psychology studies in order to satisfy a research participation component of their course grade. Because the introductory psychology courses satisfy a general education requirement for social science, students enrolled in these courses are typically representative of freshman and sophomore undergraduate students at ISU.

Information about the study was posted online on a department-managed research sign up system. The purpose of the study was described as learning about "how people evaluate video games" in general, with passing mention of virtual reality games. Participation was discouraged if certain health criteria listed in the HTC Vive Health and Safety Guide applied ("HTC Vive Health and Safety Guide," n.d.). These health criteria were also listed in the informed consent document and will be described in further detail in the following section. Participants initiated participation in the present study by contacting the experimenters and arranging a study appointment.

2.2. Materials

Informed consent document

Because it is important to minimize participant expectation of possible negative effects from virtual reality experience, the informed consent document that explained the study purpose to the participant will be described. First, health guidelines provided with the virtual reality system used for the study, the HTC Vive, were used to describe health related exclusion criteria ("HTC Vive Health and Safety Guide," n.d.). These include: pre-existing serious medical conditions (such as a heart ailment), conditions that affect the ability to safely perform physical activities, psychiatric conditions (such as

anxiety disorders or post-traumatic stress disorder), previous history of epilepsy or seizures, loss of awareness, or other symptoms linked to an epileptic condition, pregnancy, or being elderly. If any of those conditions are met, the HTC Vive Health and Safety Guide recommends consulting with a physician before using the system. The guide also recommends not using the system if you are sick, fatigued, under the influence of intoxicants/drugs, or not feeling generally well. The aforementioned criteria were listed on the consent document and potential participants were instructed to not participate if any applied to them.

The purpose of the study was stated as learning about "how people evaluate video games" in general, with only passing mention of virtual reality games to minimize potential expectation effects of the VR experience. Because virtual reality is a novel technology, if the study had been described explicitly as a "virtual reality study", participant interest could have biased toward those excited about VR and away from those that are reluctant about VR.

The study procedure was described as involving playing a video game that could include "problem solving, sports, puzzles, role playing, violence, simulations, or could be educational software on either a desktop PC or using a virtual reality system". The description of risks or discomforts included a statement that some people may experience mild motion sickness when playing video games, but that the participant should notify the experimenter immediately if they feel uncomfortable while playing the game and wish to stop. The document also stated that there were no negative consequences for stopping early. This description and phrasing was chosen purposefully to minimize expectation of a negative virtual reality experience while

balancing ethical disclosure of study procedure. The document also informed the participant that their study session would be recorded by video.

Exclusion check

Due to concern about participants not reading the consent document carefully, the exclusion criteria were also presented on a separate form, with each criterion an individual item. The participant was asked to indicate whether each exclusion item applied to them by responding either "No" or "Yes". These questions were administered via web-based form and completed on a desktop PC using a keyboard and mouse. If a participant answered "Yes" to any of the items, the experimenter informed them that they were not eligible to participate.

Virtual reality

The virtual reality system selected was the HTC Vive. The system is an accessory for a standard desktop PC, although the latest graphics processors are required for the highest resolution experience. The HTC Vive system includes a head-mounted display, two handheld controllers, and two laser emitters ("base stations") to facilitate tracking of the HMD and controllers. The HMD and controllers are motion tracked in position and rotation in high precision and 360 degrees within the "play area", the tracking bounds defined during system setup based on available space. The two base stations, mounted above the user in opposite corners of the space, emit laser light that is detected by the HMD and controllers. This laser positioning system, combined with gyroscope and accelerometers within the tracked devices, allow for sub-millimeter tracking precision within a 3.5 meter by 3.5 meter area. The HMD is connected to the computer with a 4.8 meter cable that combines data, audio, and video; the handheld

controllers connect to the PC wirelessly. The HMD weighs 555 grams and uses low-persistence OLED, with a combined resolution of 2,160 by 1,200 pixels and 90 Hz refresh rate. The effective horizontal field of view is 110 degrees. Fresnel lenses are used to reduce weight. Distance between lenses is adjustable to accommodate different interpupillary distance and the screen can be moved further from the face to accommodate glasses or user preference. The HMD has a front facing camera that is used primarily as part of a safety system. This system gives the user a way to "see" their surroundings without removing the HMD so that they can avoid physical obstacles. When the user is too close to the play area bounds, a fixed grid outlining the bounds of play area is superimposed on the virtual environment. An outline of physical objects in the environment is also shown based on the view of the HMD camera. The safety grid appears and disappears seamlessly and unobtrusively.

Virtual environments

Each participant played one VR game from a set of three games. During the first phase of the study, participants were assigned to play a specific game to facilitate the development of study protocol and procedure. Once protocols were developed and tested for each game, participants were randomly assigned to play one of the three games.

The selected games were some of the most highly reviewed and played on Steam (as of September 2016) that were also conducive to the study setting. Games that required online multiplayer interaction were excluded due to potential confounding variability of gameplay experience. Games that featured excessive violence were also excluded. Multiple games were chosen in order to increase construct and external

validity of results by testing games with a variety of characteristics. For example, games were chosen with different visual motion control mechanics and different amounts of physical movement.

Characteristics were not rigorously controlled between games which limits causal inference for between game comparisons.   However, comparisons between games should still be insightful and indicative of construct validity for the measures of interest. Descriptions of the games follow.

Waltz of the Wizard is a "wizard-simulation" experience. The game is designed for a "room-scale" experience; the player is tracked within a defined "play area" of at least 2m by 2m. The player stands at a table in a room of a tower, designed in the style of medieval fantasy. The physical play area corresponds to the area behind the virtual table. On the table are a number of magical ingredients and objects. Players add ingredients to a cauldron to gain magical powers, such as the ability to animate objects by touching them, making objects levitate, and shooting fire from their hands. The game provides little guidance and encourages players to exercise curiosity to explore their magical abilities. The game is "experiential". There are no achievements nor threats that end the game, players simply choose to end the game on their own accord. Because point of view visual motion is controlled by head movement and navigation visual motion is controlled by stepping around the play area, gameplay will result in only concordant visual motion and vestibular information.

NoLimits 2 Roller Coaster Simulation Demo is a sophisticated roller coaster simulator developed for the design of roller coasters. Users can virtually experience riding the roller coasters. The game is designed for a seated VR experience. Players

don the HMD and sit in a chair and are presented with a point of view according to being seated in the roller coaster. No controls are used once the ride starts. The environment is rendered in a photo-realistic way. The demo version of the game includes three different roller coasters, each lasting around three minutes. Pilot testing indicated the roller coasters could be ordered by degree of evocative stimulus from least to greatest. Participants "rode" the "Forest Hills Park" roller coaster first. Then the participant was asked if they would like to try the next roller coaster. If they agreed, they rode "Hybris". After "Hybris", they were asked again if they would like to try the next roller coaster. If they agreed, they rode "Wilderness Park". Although point of view visual motion is controlled by head movement, because navigation or travel (i.e., riding the roller coaster) is controlled by the virtual environment, gameplay will result in discordant visual motion and vestibular information. Although head movement results in concordant point of view visual motion, visual motion from turning and accelerating will result in visual motion information without according vestibular information.

VR Karts is a cartoon-like racing game. The game is designed for a seated VR experience. Players don the HMD and sit in a chair and are presented with a point of view according to being seated in a racing go-kart. A standard video game controller pad (e.g., the XBox One controller) is used to drive the kart, however, head motion controls point of view. This allows the player to drive in one direction and while looking in a different direction. The player races against other computer-controlled kart racers. The objective is to be the first to cross the finish line of the race track in a given number of laps. An interesting gameplay component is that players can enable "weapons" that distract and slow down opponent racers. Although point of view visual motion is

controlled by head movement, because navigation or travel (i.e., driving the car) is controlled by the virtual environment, gameplay will result in discordant visual motion and vestibular information. Although head movements result in concordant point of view visual motion, visual motion from turning and accelerating will result in visual motion information without according vestibular information.

Game overview and instructions

In order to introduce participants to virtual reality consistently, materials were created to provide an overview for each game and explain the game control mechanics. These materials were created by combining marketing and instruction materials available for the HTC Vive, Steam, and refinements of game-provided instructions. To acquaint participants with the virtual reality system, general instructions for using the VR system were provided before game specific instructions.

Acclimation video

In order to acclimate participants to the study setting, participants watched a video for 5 minutes at the beginning of the study session. The video was titled "KYOTO - CRUISE 2010 [京都] ", available on YouTube (https://www.youtube.com/v/bSyxra4iYro). The content of the video is a "street cruise" of the city of Kyoto, Japan. It is similar to a documentary of walking the streets of a city throughout the day. However, no narration is provided; the audio track consists of ambient street sounds. The video was chosen in hopes of exposing participants to low stimulation content without inducing excessive boredom.

2.3. Measures

Simulator Sickness Questionnaire

Participants completed the Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993) immediately after VR game play. See section 1.4 for a detailed description of items, response options, and scoring. The SSQ was converted to a web form. Items and response options were displayed in the same way as the paper form. Participants completed the SSQ on a desktop PC using a computer mouse to select responses. Ostensibly, cybersickness can be operationalized by SSQ scores, although establishing empirical evidence for this operationalization is the primary objective of the present study.

Virtual reality game experience evaluation

Participant evaluation of the VR game play experience is operationalized as response to a *virtual reality game evaluation* scale. After gameplay, participants rated the game on 16 dimensions. The rating items were selected to cover three hypothetical dimensions related to subjective attitudes and physiological symptoms of VR game play: enjoyment, challenge, and symptoms. The symptom items related to some symptoms assessed by the SSQ. Example items include "exciting", "challenging", "painful", and "nauseating". Each item was rated on a five-point Likert scale, from "strongly disagree" (-2) to "strongly agree" (2). The game evaluation scale items were presented with three follow-up questions: whether the participant had played the game before, whether they would play the game again (if they had the chance to in the future), and any other comments they may have about their gameplay experience. These

questions were administered via web-based form and completed on a desktop PC using a keyboard and mouse.

Physiological indicators

Measurements of physiological indicators operationalized sympathetic nervous system activity. Specifically, electrodermal activity (EDA) was measured by an Empatica E4 wristband worn on the non-dominant hand wrist for the duration of the study session. EDA is recorded by the E4 wristband in microsiemens (µS) at a sampling rate of 4 Hz.

Video game and virtual reality experience

Video game and virtual reality experience were operationalized by responses to a brief game and media habits questionnaire adapted to include items regarding virtual reality. Items cover amount of game play time over various time periods during a typical week and weekend day. The questionnaire was administered via web-based form and completed on a desktop PC using a keyboard and mouse.

Motion Sickness History

Motion sickness history was operationalized as responses to the Motion Sickness Susceptibility Questionnaire (MSSQ) (Golding, 1998). The MSSQ begins with a single item asking for a rating of susceptibility to motion sickness, and then assesses specific previous experiences with motion sickness resulting from exposure to motion-based contexts such as cars, boats, and funfair rides. Items cover both childhood experiences (before age 12) and adult experiences. A total susceptibility score for the MSSQ is computed based on the responses to the specific motion-based contexts. The

single-item susceptibility rating is scored separately. These questions were administered via web-based form and completed on a desktop PC using a keyboard and mouse.

Demographic information

A set of demographic questions assessed age, sex, and ethnicity. Two additional items asked about handedness and usage of vision correcting glasses or contacts. The ethnicity question included response options coinciding with ethnicity reports published by Iowa State University (ISU) so that the representativeness of the sample relative to the study population could be judged ("Iowa State Gender, Ethnicity & Residence Reports," n.d.). Following the design of the ISU ethnicity reports, participants that indicated they are an international student were not asked the ethnicity question.

2.4. Procedure

Upon arrival to the research laboratory, the participant was greeted by the experimenter and directed to a room prepared specifically for virtual reality research. The room was large enough to define a 4 meter by 4 meter "play area" for the HTC Vive system. Any potential obstacles were removed. Against a wall, outside of the bounds of the defined play area, was a desk with a computer monitor, the virtual reality PC computer tower, and chair. Measures designed as computer-based forms were completed at the desk using the PC and computer keyboard and mouse. The far corner of the play area was approximately 4.25 meters from the computer tower, allowing free movement within the play area without excessive tension on the HMD cable. Cameras mounted against the ceiling recorded video of the study session.

The participant was provided the informed consent document to read and indicate consent if they agreed to participate, followed by the exclusion check. After the exclusion check, the experimenter put the EDA band on the participant's non-dominant hand wrist. The experimenter then left the participant alone to watch the acclimation video for five minutes.

After the acclimation period, participants were told that they would play a virtual reality video game and were asked to read the virtual reality system overview and game instructions. To decrease experiment and experimenter demand effects, before donning the HMD, participants were instructed that they could play the game for as long as they wished, up to 20 minutes, and could stop playing at any time. Participants were left alone in the study room for all games except *NoLimits 2 Roller Coaster Simulation*. Because each roller coaster only lasted three minutes, it was more efficient to have the experimenter remain in the room with the participant and wait to either initiate each roller coaster or help the participant move on to the next part of the study. When outside of the study room, the experimenter monitored the participant through a video monitor in a control room.

Immediately following gameplay, the HMD was removed and the participant was directed to the computer to complete the SSQ, followed by the game experience evaluation and other questionnaires (video game and virtual reality experience questionnaire, motion sickness history questionnaire, and demographics). After all questionnaires were completed, the EDA band was removed and the participant was debriefed.

CHAPTER 3. RESULTS

3.1. Descriptive Statistics

Of the 202 participants, 37% were female. The median age was 19 (SD = 1, range: 18—30). The majority were White (82% White, 5% Black, 5% Two or more races, 4% Hispanic/Latino of any race, 3% Asian, 1% Native Hawaiian or Pacific Islander, 0% American Indian or Alaskan Native) and 12% were international students.

Most reported good or excellent health (54% Good, 38% Excellent, 8% Fair; recall that those in poor health were excluded from participating). They reported having their usual amount of sleep the previous night (73% Usual amount, 16% Less, 11% More). Only two participants reported taking any cold, flu, allergy, or other medications that could cause drowsiness. The majority did not wear some form of vision correction (56% No vision correction, 25% Glasses, 20% Contacts). The majority were right-handed (90%).

Motion sickness susceptibility

In response to the single item asking about susceptibility to motion sickness, the majority reported not being susceptible to motion sickness at all (46% Not at all, 35% Slightly, 13% Moderately, 7% Very much so). MSSQ total scores indicated low motion sickness susceptibility (*M* = 19.3, Median = 12.3, *SD* = 20.9, range: 0—102.3). In the original MSSQ validation study (Golding, 1998), the correlation between the single item response and MSSQ total score was .63, which was comparable to the present study (*τ* = 0.48 [0.41, 0.55]). However, the present study sample exhibited a more polarized distribution, with more participants indicating lower and higher amounts of motion sickness susceptibility. Golding (1998) reported 31% "not at all" and 3% "very much so".

The mean MSSQ reported by Golding was much higher (M = 45.5, SD = 37). Furthermore, Golding reported a difference in susceptibility between males and females, however, such a difference was not found in the present study (M_{diff} = -3.34, 95% CI: [-9.95, 3.27]; d = -0.15 95% CI: [-0.44, 0.14]).

These results may indicate that the present study sample of undergraduate students is less susceptible to motion sickness than a general population. There could be several reasons for this, but the most inferentially significant could be that participants self-selected based on the study posting description. Although the posting did not mention motion sickness and minimized advertising virtual reality games, the study was framed as an evaluation of video games in general. Perhaps participants that are interested in video games, and therefore more likely to participate in the present study, are less susceptible to motion sickness. Given the theoretical relationship between motion sickness and cybersickness, the generalizability of subsequent responses to the virtual reality experience may be limited.

Video game and VR experience

However, most participants reported not regularly playing video games, so it does not seem likely that participants were motivated to participate because the study was advertised as being about video games. For a typical week, the modal number of hours spent playing video games was zero and the distribution was highly positively skewed (M = 4.9, Median = 3.8, SD = 5, range: 0—22). Males reported slightly more hours spent playing video games than females (M_{diff} = 2.93 [1.54, 4.32], d = 0.61 [0.31, 0.9]).

The majority reported no prior experience using virtual reality (68% None at all, 28% A little, 2% A moderate amount, 1% A great deal). These results may indicate that the study sample was not self-selectively biased toward video gamers or those with prior VR experience, although it is difficult to determine without a comparative dataset. Typical weekly amount of video game play does not necessarily equate with video game interest, but the lack of video game play and VR experience reported somewhat ameliorates concerns about the wording of the study posting affecting self-selection.

Virtual reality games

Among the three games, *Waltz of the Wizard* was played by the most participants (Table 1). Only two had prior experience playing the game (*VR Karts* in both cases). The average amount of play time across all games was 17.3 (*SD* = 4.7) minutes, but varied between games, ranging from 8.9 (2.1) minutes on average for *NoLimits Roller Coaster Simulator* to 19.4 (2.1) minutes on average for *Waltz of the Wizard*. Of course, with only ten minutes of possible gameplay, it would be expected for *NoLimits Roller Coaster Simulation* to be played for the least amount of time. *VR Karts* was played slightly fewer minutes than *Waltz of the Wizard* (M_{diff} = -2.66 [-3.96, -1.36], *d* = -0.62 [-0.93, -0.31]).

Recall that the games were categorized into two different visual motion control contexts, either concordant (game visual motion corresponds with physical motion) or discordant (game visual motion does not necessarily correspond with physical motion). These were played by approximately equal number of participants (51%).

Physiological response (EDA)

Physiological indicators of sympathetic nervous system were operationalized by electrodermal activity (EDA). Unfortunately, there were challenges collecting physiological data using the Empatica E4 wristband. The E4 band used in the study consistently failed to record measurements due to a firmware problem, leading to a large loss of data. Additionally, experimenters made errors when recording event timing data, so it was not possible to identify timing events for some participants. As a result, analyzable EDA data were only available for 23 participants.

EDA is recorded by the E4 wristband in microsiemens (µS) at a sampling rate of 4 Hz. Visual inspection of time-series plots for EDA data for each participant indicated that the sensor was prone to anomalous response spikes. A 5-second rolling average smoothing window was applied to remove the spikes from each series. EDA is difficult to interpret as a raw score because of variation in physiological responsiveness and differences in skin conductivity. Range corrected EDA values were computed by dividing each observation in a series by the difference of the maximum and minimum responses for that series. Consequently, the range corrected values are bound between zero and one.

In the context of the present study, the primary metric of interest is change in EDA response over study session epochs, specifically: pre- VR game play, during VR game play, and post- VR game play. Calculating the change of EDA over those epochs for each participant time-series provided a comparative metric for interpreting change in physiological response. Therefore, for each participant time-series, linear models were fit to the (smoothed, range-corrected) EDA responses for each epoch (see Figure 8 for

an example). The slope coefficient from the linear model fit was used to summarize the change of EDA for that participant for each epoch. Because the EDA values were range corrected, the slope values computed were multiplied by 1000 to facilitate communication.

Due to low sample size for the EDA data, the empirical distributions of slopes were sparse. However, distributions of slopes for the pre-VR and VR epochs appeared symmetric (Figure 9). EDA increased over the pre-VR and VR epochs and decreased over the post-VR epoch (M = 0.3, SD = 0.4; M = 0.1, SD = 0.4; M = -0.7, SD = 0.6).

Interest in playing again

The majority of participants indicated that they would play the virtual reality game again (29% Probably yes, 25% Definitely yes, 23% Might or might not, 23% Probably not, 6% Definitely not). However, interest varied between games with the majority indicating they would play *Waltz of the Wizard* again, but a minority for *VR Karts* (Table 1 and Figure 10).

3.2. Latent Variable Measurement

Virtual Reality Game Evaluation (VRGE)

Participants rated the VR game they played on 16 items (five option ordinal scale, from "Strongly disagree" to "Strongly agree"). Empirical distributions of the responses varied between items, but there was little indication of normally distributed "partially observed" variables underlying the observed distributions (Figure 11).

Dimensionality

To evaluate the latent factor structure of the 16 VR game rating items, an Item Response Theory (IRT) analysis was conducted. Given the ordinal structure of the response items, graded response models were fit to the data (Samejima, 1969). IRT analyses were computed with the *R* package *mirt* (R Core Team, 2017; Chalmers, 2012). Inspection of polychoric correlations between items did not reveal any obviously questionable items that were too highly correlated with any other items, or too low in correlation with all other items (Table 3). Inspection of the rank order correlations (Kendall's tau) between items did not indicate any obvious issues with including all items in construction of the latent class model.

First, a confirmatory model grouping items into the three hypothetical factors was fit ("Fun", "Difficulty", and "Physiological stress"). The model was specified to allow for correlation between the three proposed factors. The confirmatory model was a poor fit to the data. See Table 4 for goodness-of-fit statistics and Table 5 for standardized and rotated (oblimin) factor loadings. For reference, the M_2 goodness-of-fit statistic is based on multi-dimensional contingency tables from the IRT model, but outperforms Pearson's χ^2 (Maydeu-Olivares & Joe, 2006).

Next, an exploratory graded response model was fit. Parallel analysis of the polychoric matrix indicated the presence of four factors (Figure 12). Again using *mirt*, an iterative process identified an exploratory four-factor graded response model. The exploratory four-factor model fit the data well (Tables 4 and 5). The four factors were interpreted as "Fun", "Nauseating", "Frustrating", and "Difficult to play". Inspection of individual item fit statistics indicated reasonable fit for all items except for "boring",

"strenuous", and "fatiguing" (Table 6). For reference, the S-χ^2 fit statistic is based on comparison of observed and expected response counts from the IRT model, but outperforms Pearson's χ^2 (Orlando & Thissen, 2000; Kang & Chen, 2011).

Information

To evaluate the information content of the model, plots were generated of item characteristic curves and test information for each factor (Figure 13), except for "Frustrating", which was composed of a single item. For the "Nauseating" factor, the "nauseating" and "dizzying" factors were the most discriminating, with "Strongly disagree" and "Strongly agree" responses clearly predicting low and high levels of the factor. The "strenuous" item was the least informative, with even high responses having moderate prediction of the factor. Overall, the model best predicted low, but positive, latent levels of "Nauseating". For the "Fun" factor, the "fun", "entertaining", and "enjoyable" items were the most informative, and "involving" and "absorbing" the least. The model best predicted lower, negative, latent levels of "Fun". For the "Difficult to play" factor, the "difficult to play" item was expectedly informative while the "challenging" item was not. Overall, the model best predicted a range of positive latent levels of "Difficult to play".

Scoring

Scoring of VR Game Evaluation responses was based on the previously described four-factor exploratory model. Component scores were computed based on unstandardized item fit coefficients. Items with standardized factor loading less than .40 were not included in the computation of that factor score. Overall, the games were rated as fun, not nauseating, not frustrating, and not difficult to play (Figure 14). The "Fun"

factor scores were based on 7 items and were negatively skewed and centered around positive ratings (M = 16.1, *Median* = 16.5, *SD* = 13.2, range: -28—33.1). The "Nauseating" factor scores were based on 6 items and were highly positively skewed and inflated at the minimum (18.8% at minimum, M = -18.2, *Median* = -24.4, *SD* = 17.4, range: -36.8—28.9). The "Frustrating" factor scores were based on 1 item and were negatively skewed and centered around negative ratings (M = -1, *Median* = -1.1, *SD* = 1.3, range: -2.3—2.3). The "Difficult to play" factor scores were based on 2 items and were positively skewed and inflated at the minimum (21.8% at minimum, M = -2.2, *Median* = -2.9, *SD* = 1.6, range: -3.9—3.5).

Relations with other variables of interest

Correlations between MSSQ scores and VRGE ratings were mixed. MSSQ and "Nauseating" were positively correlated (τ = 0.24 [0.15, 0.32]), but the correlations between MSSQ and "Fun" and "Frustrating" were inconclusive (τ = -0.09 [-0.18, 0] and 0.05 [-0.04, 0.14], respectively). The correlations between game play habits (in hours of game play per typical week) and VRGE factors were inconclusive. "Fun" ratings were slightly negatively correlated with prior VR experience (τ = -0.11 [-0.2, -0.02]), but were inconclusive for "Nauseating", "Frustrating", and "Difficult to play" (τ = -0.08 [-0.17, 0.01], -0.01 [-0.1, 0.08], -0.07 [-0.16, 0.02], respectively).

Discordant visual-motion games were rated as more nauseating than concordant (M_{diff} = 15.6 [11.18, 20.03], d = 0.98 [0.69, 1.27]). Length of game play in minutes was negatively correlated with "Nauseating" ratings (τ = -0.26 [-0.34, -0.17]), positively correlated with "Fun" ratings (τ = 0.18 [0.09, 0.27]), but was not related with "Frustrating" or "Difficult to play" ratings (τ = 0.05 [-0.04, 0.14] and 0.07 [-0.02, 0.16], respectively).

Willingness to play the game again was highly positively correlated with "Fun" ratings (τ = 0.67 [0.62, 0.72]), negatively correlated with "Nauseating" and "Frustrating" ratings (τ = -0.29 [-0.37, -0.2] and -0.17 [-0.26, -0.08], respectively), but not with "Difficult to play" ratings (τ = 0.02 [-0.07, 0.11]).

Simulator Sickness Questionnaire

Recall that responses to the 16 SSQ items were rated on a four option ordinal scale, from "None" to "Severe". Empirical distributions of responses were consistently positively skewed and inflated to the "None" response option (Figure 15, see Table 7 for response counts). There was little indication of normally distributed "partially observed" variables underlying the observations. In fact, visual inspection would seem to imply clearly that the virtual reality games did not evoke significant manifestation of SSQ symptoms.

Dimensionality

Given the ordinal structure of the response items, combined with the "zero-inflation" and high positive skew observed in the present sample, it would not seem reasonable to attempt to fit standard factor analytic models. To evaluate the latent factor structure of the 16 SSQ items, an Item Response Theory (IRT) analysis was conducted, primarily with the intention to test the latent model proposed by Kennedy et al. (1993). Although an IRT analysis is very different from a standard factor analysis, the objective of identifying latent structure is the same. Therefore, results of the IRT analysis are conceptually comparable to standard factor analysis.

Inspection of rank order correlations between items did not reveal obviously questionable items that were too highly correlated with any other items, or too low in

correlation with all other items (Table 8). Although inadequate for inference, the rank order correlations did not indicate any obvious issues with including all items in construction of the latent trait model.

A confirmatory graded response model was fit. Because the factor analytic model proposed by Kennedy et al. (1993) used orthogonal rotation, the model was not specified with correlation between the three factors. The confirmatory model was a poor fit to the data ($M_2(70) = 447.4$, $p = 0$; RMSEA = 0.167 [0.152, 0.182]; TLI = 0.795). See Table 9 for goodness-of-fit statistics for all SSQ models tested and Table 10 for standardized and rotated (oblimin) factor loadings.

Having failed to find support for the Kennedy et al. (1993) factor structure, an exploratory graded response model was then fit. Parallel analysis indicated the presence of seven factors. Because of the challenges of interpreting seven factors among 16 items, and the sparsity of the data, it was decided to fit a four-factor graded response model. The exploratory four-factor graded response model was a modest fit to the data (Table 8). The four factors indicated could be described as "Nausea", "Difficulty concentrating", "Dizziness", and "Fatigue". However, the "headache" and "eyestrain" items did not load to any factor. Based on anecdotal evidence, this was counterintuitive as some participants verbalized experiencing eyestrain to experimenters.

An alternative measure of cybersickness (CSQ)

The counterintuitive and poor fit of the four-factor exploratory model for SSQ responses 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. Although most participants did not report experiencing symptoms, some did, as

indicated by response to those SSQ items that clearly correspond to feeling unwell. Therefore, it was decided to try identifying a possible measure of cybersickness by testing a selection of SSQ items that would appear to clearly indicate sickness. For example, although a person could feel sweaty and fatigued as a result of cybersickness, the physical nature of playing VR games would also be expected to cause sweating and fatigue without sickness. The objective of reducing the items included in subsequent exploratory models was to identify those items that any positive response would indicate physiological distress. This approach is conceptually similar to the development of the SSQ from the MSQ. For brevity, this potential new measure, based on SSQ items, will be referred to as the CyberSickness Questionnaire (CSQ).

Dimensionality

Review of the 16 SSQ items indicated seven that clearly indicated sickness: headache, eyestrain, nausea, blurred vision, dizzy (eyes open), dizzy (eyes closed), and vertigo. To address the sparsity of fitting graded response models based on seven items and four response options, it was decided to amalgamate the "Severe" and "Moderate" responses. Parallel analysis indicated the presence of three factors, however, a three-factor model had too few degrees of freedom to compute the M_2 goodness-of-fit statistic. A two-factor exploratory graded response model was a good fit to the data, but had only one degree of freedom (Table 9).

In order to increase degrees of freedom, more items would need to be included. Among the remaining SSQ items, "difficulty focusing" and "fullness of head" were chosen as the next items with the clearest indication of physiological distress. "Difficulty focusing" is similar to "blurred vision", and anecdotally some participants reported

feeling excess pressure from wearing the head-mounted display. Responses were then amalgamated for the two additional items.

Parallel analysis of the nine items indicated presence of three factors. An exploratory three-factor graded response model, based on the nine items with three response options, was a good fit to the data (Table 9). However, the model had only three degrees of freedom, which significantly hinders interpretability and application to future data sets. Given the sparsity of the data, it was decided to fit a two-factor model.

The exploratory two-factor graded response model, again based on the nine items with three response options, was a very good fit to the data ($M_2(10) = 4.68$, $p = 0.912$; RMSEA = 0 [0, 0.03]; TLI = 1.016). The two factors were interpreted as "Dizziness" and "Difficulty focusing." Item fit was good, except for "blurred vision" (see Table 11). Inspection of individual item fit statistics indicated reasonable fit for all items except for "blurred vision" (Table 12).

Information

To evaluate the information content of the model, plots were generated of item characteristic curves and test information for each factor (Figure 16). For the "Dizziness" factor, the "dizziness eyes open" and "dizziness eyes closed" items were highly informative (as expected), followed by "vertigo". The "nauseous" item was the least informative. Overall, the model best predicted a narrow range of low, positive, levels of latent "Dizziness". For the "Difficulty focusing" factor, item information was more difficult to interpret. All items described a moderate amount of information, with some items working better for extreme responses, such as "blurred vision" and others better for

middling, such as "eyestrain". Overall, the model best predicted a broader range of positive levels of latent "Difficulty focusing".

Scoring

Scoring of CSQ responses was based on the previously described two-factor exploratory model of nine items with amalgamated response options. Items with standardized factor loading less than .40 were not included in the computation of that factor score. Overall, participants did not report feeling dizzy or having difficulty focusing after VR game play (Figure 17). The "Dizziness" factor scores were based on 5 items and were highly zero-inflated and positively skewed (57% zero values; *M* = 7.1, *Median* = 0, *SD* = 12.7, range: 0—46). The "Difficulty focusing" factor scores were based on 4 items and were highly zero-inflated and positively skewed (40% zero values; *M* = 2, *Median* = 0.8, *SD* = 2.5, range: 0—9.5).

Relations with other variables of interest

To the extent that some component of cybersickness is based on visual-motion and vestibular information discordance (conceptually the VIMS component), it would be expected that cybersickness would positively correlate with motion sickness susceptibility. Accordingly, MSSQ was positively, although modestly, correlated with both "Dizziness" (τ = 0.26 [0.17, 0.34]) and "Difficulty focusing" (τ = 0.22 [0.13, 0.31]). Video game experience was slightly negatively correlated with "Dizziness" (τ = -0.14 [-0.23, -0.05]), but the correlation with "Difficulty focusing" was inconclusive (τ = -0.07 [-0.16, 0.02]). The results suggested that prior VR experience was not correlated with either "Dizziness" or "Difficulty focusing" (τ = -0.01 [-0.1, 0.08] and -0.04 [-0.13, 0.05], respectively).

Games with greater visual-motion discordance should cause greater cybersickness. Accordingly, participants that played a discordant visual-motion game had higher CSQ scores for both factors, "Dizziness" (M_{diff} = 7.38 [3.88, 10.88], d = 0.59 [0.3, 0.87]) and "Difficulty focusing" (M_{diff} = 0.94 [0.24, 1.64], d = 0.37 [0.09, 0.65]).

Experiencing cybersickness should discourage continuation of VR. Accordingly, minutes of game play was negatively correlated with "Dizziness" (τ = -0.33 [-0.41, -0.25]). However, the results were inconclusive as to the correlation of minutes of gameplay and "Difficulty focusing" (τ = -0.07 [-0.16, 0.02]). The correlation is interpreted as inconclusive because the 95 percent confidence interval for the estimate is too wide and close to (or includes) zero.

If a virtual reality experience causes cybersickness, it would be expected to be rated as less fun. Accordingly, the "Fun" factor of the VR Game Evaluation was negatively correlated with both "Dizziness" (τ = -0.27 [-0.35, -0.18]) and "Difficulty focusing" (τ = -0.25 [-0.33, -0.16]). Similarly, cybersickness should decrease willingness to experience virtual reality again. Accordingly, willingness to play the VR game again was negatively correlated with both "Dizziness" (τ = -0.3 [-0.38, -0.21]) and "Difficulty focusing" (τ = -0.27 [-0.35, -0.18]). Figure 18 depicts the relation between willingness to play again and VRGE "Fun" rating, sized by each CSQ factor score.

3.3. Research Questions

Having described the sample data and established scores for measures of the latent variables of interest, specific research objectives can be addressed. Recall that the the main research objectives were to: i) psychometrically evaluate the Simulator Sickness Questionnaire in the context of virtual reality; ii) evaluate the potential of

alternative measure of cybersickness, potentially based on SSQ items or report of attitudes about the VR experience; iii) evaluate a new measure of subjective evaluation and attitudes of the VR experience, and explore general subjective responses to state-of-the-art, consumer-oriented, VR systems and the potential of physiological measures as indicators of cybersickness. These objectives have been rephrased below for additional clarity.

In general terms, the analytic portion of psychometric evaluation is a process of identifying scores for latent variables based on observations and then evaluating the reliability and validity of those scores. Scores for latent variables can be determined after identifying the dimensionality and applicable model structure. It should be noted that the concept of "reliability", and related statistics that most social scientists are familiar with are based on Classical Test Theory. The corollary to reliability from the Item Response Theory perspective is "information." Validity of a measure is always difficult to establish and can be addressed in many ways. In the following analysis, evidence of concurrent validity can be established by evaluating relations between scores of latent variables with other variables of interest.

Is the SSQ a valid measure of cybersickness?

The dimensionality analysis of the SSQ began with a confirmatory graded response model based on the structure reported by Kennedy et al. (1993). Goodness-of-fit statistics indicated that this model was a poor fit to the data ($M_2(70) = 447.4$, $p = 0$; RMSEA = 0.167 [0.152, 0.182]; TLI = 0.795). Therefore, it can be inferred that the Kennedy et al. SSQ factor model does not apply to responses collected in the context of virtual reality. Parallel analysis indicated the presence of seven factors, which seemed

implausible among 16 items. An exploratory analysis attempting to fit a four-factor graded response model was unsuccessful. Without a model that determines how item scores relate to the latent variable of interest, cybersickness, the SSQ cannot be considered a valid measure of cybersickness.

Is the CSQ a valid measure of cybersickness?

The exploratory two-factor latent trait model for the CSQ fit the data very well. Analyses indicated that the items were generally informative, although less so for "Dizziness" than "Difficulty focusing". Overall, there was supporting evidence for the concurrent validity of the CSQ as a measure of cybersickness based on relations of CSQ scores with other variables of interest. The question of content validity is more difficult to address. Given the lack of strong theory about which symptoms should be expected to underlie cybersickness, one approach to evaluating content validity would be to refer to the individual item fit. All but one item exhibited satisfactory fit to the model, yet the item that did not, "blurred vision", would seem intuitively to be an important indicator of potential negative effects of accommodation from head-mounted display design. Perhaps the presence of both the "difficulty focusing" and "blurred vision" items created confusion for some participants and lead to unpredictable response patterns.

On the whole, there does seem to be preliminary evidence to support the validity of the CSQ as an informative measure of cybersickness. Interest in playing the VR game again is one of the most interpretable indicators of how participants felt about their VR experience. Although CSQ factors did not strongly predict interest to play again, Figure 18 demonstrates how cybersickness, operationalized by CSQ scores,

may interact with "Fun" (as operationalized by VRGE). Most participants reported having fun, and most indicated interest in playing again. Nonetheless, Figure 18 indicates that a fun experience and interest in playing could coincide with high levels of cybersickness. At the same time, those that did not have fun and had less interest in playing again seem to have experienced relatively more cybersickness.

Is the VRGE a valid measure of subjective response to VR game play?

The dimensionality analysis of the VR Game Evaluation (VRGE) indicated that the hypothesized three-factor graded response model was not supported by the data. An exploratory four-factor model was a modest fit, with one factor loaded to a single item ("frustrating"), and another factor, "Difficult to play", was composed of only two items. Individual item fit was satisfactory overall, however, there is some indication that the model does not capture negative physiological influences of VR game play because of the poor fit of the "strenuous" and "fatiguing" items.

The "Nauseating" factor would be considered the component that measures cybersickness, but in comparison to the CSQ results, it is clear that the VRGE does not cover vision symptoms adequately. It was encouraging to find a distinction between the attitudinal components of "fun", "frustrating", and "difficult to play". However, inspection of the open-ended comments indicated that the degree of "immersion" was often mentioned by participants, but was not addressed by the VRGE.

Evidence of concurrent validity were generally mixed and somewhat conflicting with the model evaluation. "Nauseating" VRGE ratings were positively correlated with MSSQ scores comparably to CSQ factors. The degree of negative correlation between "Nauseating" ratings and game play in minutes, as well as willingness to play the game

again, were also comparable to CSQ "Dizziness". The difference in "Nauseating" ratings between discordant and concordant games was larger than the difference in feelings of "Dizziness" from the CSQ. Thus, the VRGE "Nauseating" ratings concur with the CSQ given the same plausible correlates of cybersickness. The VRGE "Fun" rating also had the strongest correlation with willingness to play again among any of the latent measures of interest, including the CSQ. However, the VRGE "Frustrating" and "Difficult to play" ratings were not as informative, with low or inconclusive correlations with other variables of interest.

Overall, there is evidence that the VRGE could become a valid measure of subjective response to virtual reality game play with further refinement. Evaluation of the model indicated possibly weak evaluation of the "Nauseating" factor and consideration of the content validity questioned the lack of items addressing vision problems, but concurrent validity of the "Nauseating" factor was supported to a similar degree as the CSQ.

What best predicts cybersickness?

Modeling the incidence of cybersickness, as operationalized by CSQ and the VRGE, as a function of other variables of interest extends understanding of the validity of the measures. It was decided to focus the scope of an exploratory predictive modeling analysis on variables that could be assessed prior to VR experience to predict CSQ "Dizziness", CSQ "Difficuly focusing", and VRGE "Nauseating". Seven candidate predictors were selected based on the available data: gender, use of vision correction, susceptibility to motion sickness (single-item), MSSQ total score, video game experience, prior VR experience, and the visual-motion concordance of the game.

There was sparse missingness among these variables, with the greatest being 13 missing MSSQ total scores due to failure to respond to a single item within the MSSQ scale. Missingness was addressed by assigning mean values for interval variables and highest frequency for nominal. Dichotomous indicator variables were created for nominal variables for the sake of interpretability. Because of low response frequency for options higher on the scale, VR experience was dichotomized to indicate either did (1) or did not (0) have prior VR experience.

Because of zero-inflation, it was decided to dichotomize each of the cybersickness-related outcome variables of interest to indicate either a zero score for that factor (0) or a positive score (1). Although the dichotomization removes some available information of the degree of cybersickness experienced, the high zero-inflation and low frequency of positive response create challenges for achieving satisfactory model fit. Pragmatically, at this stage of the development process for a measure of cybersickness, it is reasonable to broadly address detection and prediction of the phenomenon. Fortunately, the breakpoint of zero is intuitive in this context, as the latent measures for cybersickness were constructed in a way to clearly indicate distress with positive response.

Inspection of correlations between the variables of interest (Table 13) revealed a potentially confounding positive correlation between the single susceptibility to motion sickness item and the game visual motion ($r = 0.17$ [0.08, 0.26]). The intention of the single susceptibility to motion sickness item is for the respondent to summarize how susceptible to motion sickness they feel that they are in general, which should mean that responses to this item could be considered to measure trait susceptibility. Because

the games were randomly assigned, there should not be a significant relation between the visual-motion of the game that was played and susceptibility. The result indicates that either playing a game with discordant visual-motion caused an increase in perception of general susceptibility to motion sickness or that randomization failed. It seems likely, therefore, that the susceptibility item did not clearly measure trait susceptibility, so it was dropped from further analysis.

Because the immediate analytical goal is prediction of cybersickness, the dataset was randomly divided into "training" and "testing" subsets. Models were identified based on the training subset, then accuracy was evaluated by fitting to the testing subset. After creation of indicator variables and removal of the single motion sickness susceptibility item, there were eight predictor variables. An exhaustive model selection process was used to identify the linear combination of these predictors that best predicted each of the (dichotomous) cybersickness outcome variables. A logistic regression model was fit to each linear combination of predictors with the *R glm* function and binomial family (logit link function). Candidate models were limited to at most five predictors. For each outcome of interest, the best fit was determined to be the model with the lowest AIC among the 218 candidate models. See Table 14 for summary statistics for the five models with lowest AIC for each outcome and Tables 15, 16, and 17 for summary information for each chosen model. The chosen model was then used to predict the probability of outcome for the test data subset. To evaluate the accuracy of the model, the prediction probabilities were rounded to zero or one. The model prediction accuracy was then described by the percentage of predictions that agreed with actual responses in the test data subset.

For CSQ Dizziness, the model with lowest AIC included three predictors: the indicator for normal vision, MSSQ score, and the indicator for discordant game motion. The model fit the training data well (Hosmer-Lemeshow $\chi^2(8)$ = 8.95, p = 0.347). Given the same normal vision status and MSSQ scores, the odds of experiencing Dizziness were 8.64 [3.38, 24.52] higher for discordant games. Given the same normal vision status and VR game visual-motion, for each unit increase of MSSQ score, the odds of experiencing Dizziness increased by 1.04 [1.02, 1.07]. The change in odds for those with normal vision was inconclusive (odds = 0.47 [0.17, 1.24]). Predictions for the testing data were reasonably accurate; 71% of the predicted outcomes aligned with the actual outcomes of the testing data.

For CSQ Difficulty focusing, the model with lowest AIC included three predictors: the indicator for wearing glasses, video game experience, and the indicator for discordant game motion. The model fit the training data well (Hosmer-Lemeshow $\chi^2(8)$ = 9.68, p = 0.288). Given the same glasses wearing status and video game experience, the odds of experiencing Difficulty focusing were 3.42 [1.41, 8.82] higher for discordant games. Given the same video game experience and VR game visual-motion, the odds of experiencing Difficulty focusing were 4 [1.36, 13.91] higher for those that wore glasses. The change in odds for video game experience was inconclusive (odds = 1.07 [0.98, 1.18]). Predictions for the testing data were modestly accurate; 60% of the predicted outcomes aligned with the actual outcomes of the testing data.

For VRGE Nauseating, the model with lowest AIC included two predictors: video game experience and the indicator for discordant game motion. The model fit the training data well (Hosmer-Lemeshow $\chi^2(8)$ = 4.87, p = 0.772). Given the same video game

experience, the odds of rating the game as Nauseating were 24.38 [4.62, 451.78] higher for discordant games. The change in odds for video game experience was inconclusive (odds = 0.89 [0.75, 1.03]). Predictions for the testing data were reasonably accurate; 82% of the predicted outcomes aligned with the actual outcomes of the testing data.

CHAPTER 4. DISCUSSION

4.1. Summary

In the present study, participants played a virtual reality (VR) videogame and then completed questionnaires about their VR game play experience, specifically in relation to cybersickness, and additional measures to assess potentially related covariates. The primary objective was to psychometrically evaluate responses to the Simulator Sickness Questionnaire (SSQ) as a measure of cybersickness. The SSQ is commonly used as a measure of cybersickness (Rebenitsch and Owen 2016); however, the measure was developed and validated for use in the context of flight simulation for military pilots (Kennedy et al. 1993), a very different context and population from typical modern VR systems and users. Further study objectives included evaluating the potential of a new alternative measure of cybersickness and subjective response to newly available consumer-oriented VR game systems.

Three different VR games were chosen as stimuli that differed in degree of visual-motion and vestibular information concordance. Based on sensory conflict theory, it was expected that playing a VR game with only concordant visual-motion information, that is, where visual-motion displayed in VR always aligned with vestibular information, would cause less cybersickness than playing a VR game with discordant visual-motion and vestibular information, where visual-motion displayed in VR did not align with vestibular information. The concordant visual-motion VR game was a "wizard simulator", where the player explored a magical workshop. The other two games, a kart racing game and a roller coaster simulation, incorporated discordant visual-motion.

Results of the present study indicated that the SSQ, at least in standard form and scoring structure based on Kennedy et al. (1993), is not a valid measure of cybersickness. Further exploratory analysis of SSQ responses did not identify a latent model that was a satisfactory fit to the data. However, there was evidence that an alternative measure, constructed by selecting specific items from the SSQ and collapsing the "Moderate" and "Severe" response options, may be a valid measure of cybersickness. This alternative measure was created to focus on symptoms of cybersickness, referred to as the CyberSickness Questionnaire (CSQ) to differentiate it from the standard SSQ items and scoring procedure. Two factors were identified among the nine items of the CSQ: "Dizziness" and "Difficulty focusing".

A component of a new measure of subjective attitudes about the VR game play experience, the VR Game Evaluation (VRGE), was also tested and found to need further refinement. The "Nauseating" factor of the VRGE addressed the degree that the VR game would manifest nausea and dizziness, but did not address vision effects that were evident from inspection of SSQ responses. The "Fun" factor of the VRGE measured positive attitudes about VR game play, but the measure did not effectively address negative attitudes.

Overall, incidence of cybersickness was very low and mild in magnitude in this sample of undergraduates playing a commercial VR game for up to 20 minutes. Based on CSQ factor scores, 57% of participants reported no dizziness at all and 40% reported no difficulty focusing. A small percentage did report experiencing "moderate" and "severe" symptoms. Logistic regression models were constructed to identify predictors of cybersickness. In support of expectations, the visual and vestibular

concordance of the visual-motion used in the game was the strongest predictor of cybersickness, with VR games that incorporated discordant visual-motion information being much more likely to evoke cybersickness than concordant visual-motion VR games. Results suggested that history of motion sickness and wearing glasses also increased the odds of evoking cybersickness. Video game play experience was also indicated as a predictor of cybersickness, although the magnitude and direction of the relation was inconclusive.

In general, the VR games were rated as being fun and participants expressed strong interest in playing again. VRGE "Fun" rating was highly correlated with interest in playing again, much more than the correlations between any of the cybersickness measures. Notably, some participants indicated strong interest in playing again and rated the game as fun even after reporting high levels of cybersickness.

4.2. Conclusions

The SSQ should not be used to measure cybersickness

Based on the results of this study, the SSQ should not be used as a measure of cybersickness. As discussed previously, there were several reasons to anticipate this, including: the relevance of the items to cybersickness based on the differences between flight simulator and VR contexts, the unrepresentativeness of the SSQ validation sample, and the questionable applicability of the particular factor analysis that was used. Others have also previously made similar arguments against the use of the SSQ in the context of VR (Rebenitsch & Owen 2016; Ames et al. 2005; Young et al. 2006; Kim et al. 2004; Bouchard et al. 2007; Bruck & Watters 2011). However, virtual reality researchers in general have not been convinced by the arguments against the SSQ and

have continued to use it to measure cybersickness. The present study has corroborated the conceptual arguments against the SSQ with strong empirical evidence.

One of the primary advantages of the present study is that of generalizability. Responses were collected from a relatively large sample for this research domain, using VR equipment that is now widely available and virtual environments that are popular for the system. Although the representativeness of the undergraduate sample to the general population of potential VR users requires further examination, the sample should be representative of the typical research population of participants in VR studies conducted in academic settings.

How to use the CSQ as a measure of cybersickness

A fortunate result of the present study was that a derivative of the SSQ, referred to here as the CSQ, appears to be a valid measure of cybersickness. Although the CSQ is based on a different factor structure and scoring approach, researchers could continue to administer the SSQ and then score it as the CSQ. This minimizes the risk of abandoning what is considered to be an established measure.

Scoring of the CSQ requires accepting a compromise of the present study. The CSQ results reported were computed based on unstandardized factor loadings from the exploratory graded response model. These component scores make full use of the available data and align with the item response theory perspective of weighting items individually to capture individual item contribution. However, because the factor loadings are estimates based on this particular sample, they have uncertainty that complicates application to future data sets. As a pragmatic compromise, it is recommended to score CSQ factors by weighting responses by the standardized factor

loadings reported in Table 11. Note that items without a number reported for that factor should not be included in the scoring of that factor.

There is clearly further work to be done to refine the CSQ. Nonetheless, based on currently available information, this is the recommended application:

- Administer the SSQ after exposure to VR.

- From the collected data, remove the General discomfort, Fatigue, Increased salivation, Sweating, Difficulty concentrating, Stomach awareness, and Burping items.

- Amalgamate responses by combining "Moderate" and "Severe" options for each item (recode "Severe" responses to "Moderate").

- Score the item responses as "None" (0), "Slight" (1), and "Moderate" (2).

- Compute scores for the two CSQ factors by taking the sum of the weighted item responses. That is, the sum after multiplying each item score by the according standardized factor loading for that item in Table 11.

Based on the suggested standardized loading scoring approach, factor scores could range from 0 to 7.52 for "Dizziness" and 0 to 5.66 for "Difficulty focusing". For reference, in the present sample, mean "Dizziness" was 1.2 (*Median* = 0, *SD* = 1.8, range: 0—7) and mean "Difficulty focusing" was 1.2 (*Median* = 0.6, *SD* = 1.5, range: 0—5.66).

Please note that there is not a "total score" for the CSQ as there is for the SSQ. This may cause some difficulty for those familiar with the use of the SSQ as the SSQ Total score is most often reported. The results of the IRT analysis indicate a two-factor model fits the present data well, however, an argument could be made that there is a

general latent factor, such as in the case of a bifactor model. If responses to the CSQ were best explained by a bifactor model, then a general factor, cybersickness, influences item responses in addition to the two identified factors, which would be interpreted as subfactors. Similarly, it could be that the two CSQ factors indicate different levels of severity of the general factor of cybersickness. That is to say, the "Difficulty focusing" factor may describe generally low levels of cybersickness and "Dizziness" describes high levels.

Evidence against these interpretations can be found by referring to a scatterplot of scores for each CSQ factor (Figure 19). If the two factors were in fact subfactors of a general factor, then a relationship between each factor score would be evident. Instead, the scatterplot indicates a nearly even dispersion of combinations of scores across the range of each factor.

Cybersickness may not be as prevalent Visually-Induced Motion Sickness

Across both measures, incidence of cybersickness was very low and overall ratings of the experience were positive. This contrasts with prior research of Visually-Induced Motion Sickness (VIMS) and simulator sickness. An important consideration is that actual VR users in practice are likely even less susceptible to cybersickness than what was observed in the present study. Given the steps taken to avoid participant self-selection bias for interest in VR, or lack thereof, the real population of VR users who purchase a VR system for entertainment or other uses, likely experiences less cybersickness than observed in the present study. This is hinted at by virtue of the fact that greater video game play predicted lower cybersickness in this study. Furthermore, the proportion of discordant visual-motion VR games administered was far greater than

the proportion of discordant visual-motion VR games that are popular among VR gamers, which further decreases the general likelihood of cybersickness. In fact, it was somewhat challenging to find discordant visual-motion games as it seems that game developers must have quickly realized that visual-motion discordance would be more likely to lead to cybersickness.

Although low incidence of cybersickness is a good thing in practical terms, there were participants that experienced significant sickness, as evidenced by the wide range of cybersickness scores and raw item responses. Fortunately, no one became ill enough that they had to discontinue the study, but experimenters noted that some were relieved to stop playing the VR game. One scientific challenge is that the group that would benefit most from further understanding of the mechanisms of cybersickness is the minority. It will be critical to either employ large samples or to identify trait predictors that predispose one to cybersickness to increase the efficiency of data collection. The results of the present study indicate that a reasonable strategy would be to recruit participants that play video games less and self-report being more susceptible to motion sickness. Although such an approach complicates data collection, it would seem better than trying to study the effect of a treatment where half the sample did not experience any cybersickness to begin with.

It is important to assess the broader VR experience

As discussed previously, it is common for VIMS and cybersickness researchers to focus on the negative response to the exclusion of the positive. In the present study, in addition to measurement of negative symptoms, a new measure was created to assess attitudes about the VR experience. These complementary measures enhanced

understanding of cybersickness. A priori, it would seem that any cybersickness would coincide with a negative impression of the VR experience; the results, however, indicated otherwise. As with most psychological phenomena, the subjective VR experience is complicated by expectations, motivations, and past experiences that focus the moment in a particular way for that particular person. Combined with the limitations of accuracy and precision of subjective attitude assessment, the problems with trying to isolate one aspect of the experience seem daunting, particularly when the excluded dimension could potentially be a principal motivator. Particularly for modern consumer-oriented VR systems, users are motivated by the expectation of VR being fun. A clear risk to VR research of ignoring the broader aspects of the VR experience is that treatment approaches that decrease cybersickness may also decrease fun, immersion, or other positive aspects.

Assessing other dimensions of VR experience also has the advantage of providing a different response frame. For example, the SSQ asks how nauseous the respondent feels "right now", whereas the VRGE asks to rate how "nauseating" the VR game was. When administered following VR, a respondent could correctly state that they are not experiencing any nausea, but that the game was nauseating; perhaps they initially experienced nausea but then either adapted or adopted a strategy to minimize it.

The issue of users independently addressing their cybersickness is worth mention as it advances argument for broader VR evaluation. Although a quantitative assessment was not feasible in the present study, it was observed that participants who seemed to experience cybersickness minimized head motion and other movement. In some cases, they may have even opted to close their eyes to decrease discordant

visual-motion information. Consider the challenge of a unidimensional measure quantitatively assessing the hypothetical qualitative response "The game was very fun, but made me feel sick until I figured out not to look up or down!"

4.3. Limitations and Future Directions

Low sample size

The sample size affects the validity of the graded response models in particular, but is related to the issue of participant recruitment. Prior to data collection, the needed sample size was determined by considering the power necessary to test standard factor analyses of the SSQ responses. Only after data collection was completed did we realize that the responses were extremely non-normal and that a standard factor analysis approach was not applicable.

The Item Response Theory (IRT) approach has many advantages for the context of the present study, but there are many open questions about the application of IRT, particularly for multidimensional, polytomous, graded response models with extremely skewed item distributions. Simulation studies of the robustness of such models have used much larger sample sizes. For example, the smallest sample size tested by Maydeu-Olivares, Cai, and Hernández (2011) was 500. Unfortunately, there are not clear recommendations for minimum sample size, but a rough idea can be found by considering the conceptual approach of IRT.

There need to be an adequate number of observations for each response option for each item, which introduces a confounding problem in the context of cybersickness. In the present data set, the extreme high response options ("Moderate" and "Severe") for CSQ items were so underrepresented that amalgamation was necessary. However,

these are the responses of greatest interest as they indicate the most severe negative effects. Conversely, the skewed distributions of the present data facilitate great estimation of measurement for those of least concern (those responding "None" and "Slight"). This limitation can be addressed in future studies by either recruiting from a general participant pool and increasing the sample substantially, or by targeting participant recruitment to those most susceptible.

There may also be legitimate concern about the representativeness of the sample to the general population. The sample was younger, more male, less ethnically diverse, and less susceptible to motion sickness than a general adult population. From an applied perspective, the sample likely had less video game play experience than the target market for current consumer-oriented VR systems. Even though the sample is likely representative of an academic research participant population, it will be important for future studies to consider how the characteristics of their research participant populations may affect the validity of the proposed measures. Validation of the measures with different populations would be recommended whenever feasible.

Incomplete evaluation of predictors

Targeted recruitment of susceptible participants requires knowledge of the core aspects of susceptibility. The present study has provided some guidance as to potential predictors, but the accuracy of the predictive models could be improved. Vision was indicated as a predictor of both CSQ factors, with normal vision possibly decreasing Dizziness and wearing eye glasses increasing Difficulty focusing (while controlling for other predictors). This result may be confounded with how well the HTC Vive HMD could accommodate particular glasses. The HMD is designed to wear with glasses, but

is dependent on the size of the glasses to fit comfortably and correctly within the display. It was noted by experimenters that some glasses wearing participants seemed to force the HMD on with their glasses in ways that seemed uncomfortable. For a couple of participants, they elected to not wear their glasses while using the HMD, but stated that their vision wasn't affected. Further evaluation of how glasses could affect wearing the HMD is warranted.

Relatedly, and as discussed previously, there are a number of ways that wearing the HMD could cause vision discomfort. The HTC Vive HMD is essentially a one-size-fits-all approach to a very biophysically diverse application and incorporates many compromises. Future studies should consider detailed collection of biophysical characteristics that could predict susceptibility, although this could make participant recruitment challenging.

Video game play experience was also indicated as a predictor of cybersickness, but with too much uncertainty to feel confident in the relation. The measure of video game play experience only addressed hours of game play during a typical week. Perhaps the type of game played is also important as visual-motion cues vary greatly between game genres. It is possible that participants with equivalent frequency of game play could differ in cybersickness susceptibility if one played first-person shooters compared to text-based strategy games. Future studies should assess both video game play frequency and game title to try to identify the relation of game type with cybersickness susceptibility.

Limited stimulus sampling

The results clearly indicated the effect of discordant visual-motion on cybersickness. This characteristic of the VR games, however, was not controlled. Because only one concordant and two discordant VR games were used, it is possible that other differences beyond concordance between games caused the difference in cybersickness. Ideally, the concordant and discordant games would be identical in every other way. For example, a game could provide a setting to switch between a teleport navigation (concordant motion) or a "walking" navigation (discordant). Although no games are currently available with this design, this is technically quite feasible. However, the ideal visual-motion control game would have less generalizability. Therefore, future studies should include both a larger number of both concordant and discordant games and try to identify an ideal visual-motion control game. Although not previously mentioned, it should be noted that several games were tested for inclusion in the present study, but failed to run consistently in pilot testing. VR is still a new technology and technical problems are to be expected. It is recommended that researchers pre-test candidate games for the full time period they intend for use in the study.

Incomplete application of the CSQ

Although the recommendation here is to use the CSQ in lieu of the SSQ, in practice the SSQ should still be administered until further testing can establish whether the CSQ model applies when only the nine CSQ items are used. Although it may be unlikely that response patterns would change significantly when administered with only nine items, it is certainly possible. The presentation of items can influence

interpretation. The nine CSQ items were selected specifically because they communicate clear negative symptoms. Perhaps in the absence of "Stomach awareness" or "Increased salivation" more participants would respond that they were experiencing nausea. It also seems possible that the "Difficulty focusing" item could be misunderstood without the "Difficulty concentrating" item. "Difficulty concentrating" would be considered to refer to mental concentration and "Difficulty focusing" to vision. When presented without "Difficulty concentrating", "Difficulty focusing" might be interpreted by some as referring to concentration. Future studies should test administering only the nine CSQ items and rewording the latter item to "Difficulty focusing vision".

Improvement of IRT analysis

As discussed previously, there are open questions regarding the application of IRT models to measures like the CSQ and VRGE that are multidimensional and polytomous. The relatively low sample size of the present study, combined with zero-inflation and high positive skew of empirical distributions, as well as underrepresentation of some response options further challenges the application chosen modeling approach. Future studies should consider alternative approaches, for example, nonparametric modeling that does not require assumptions about the distributions of the latent variables.

Collection of physiological response

As reported, the collection of electrodermal activity (EDA) ended soon after the study began due to technical problems with the Empatica E4 sensor wristband. The available data were qualitatively interesting and showed promise for future study, but

were inferentially inadequate. Although it was not reported here, pilot tests were conducted to record three-dimensional coordinate position of the HMD using the high fidelity of the HTC Vive VR system with the intention of possibly detecting instability through head position variability. Unfortunately, software bugs in the position recording program caused the VR system to stop working and it was decided to limit technical complication in the present study and not include the head position recording program in the procedure. Capturing quality physiological responses remains an objective for future research and will certainly enhance understanding of concurrent validity of cybersickness measures.

4.4. Closing Statement

The present study provides convincing data to argue against the rote application of the SSQ to cybersickness. Exposure to the challenges of psychological measurement is not uniform across the social sciences, and seems even less so in the engineering and computer science fields that have lead the development of VR systems. The phenomenon of cybersickness presents particular measurement challenges that require further consideration, and in some ways pushes the limits of current psychometric methodology. This is a worthy challenge, as interest in virtual reality technology has only continued to grow since this research project began. Virtual reality has the potential to be a transformative communication and experiential technology. Hopefully the present line of research can contribute to making it as accessible and inclusive as possible.

REFERENCES

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. *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, *82*(3), 168–176. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15767873

Anthes, C., García Hernandez, R., Wiedemann, M., Kranzlmüller, D. (2016). State of the Art of Virtual Reality Technologies. In *2016 IEEE Aerospace Conference*. Big Sky, Montana, United States.

Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation*, *19*(5). Retrieved from http://pareonline.net/getvn.asp?v=19&n=5

Barrett, J. (2004). *Side effects of virtual environments: A review of the literature*. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA426 109

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika*, *65*(1), 23–28. https://doi.org/10.1007/bf02294183

Bouchard, S., Robillard, G., & Renaud, P. (2007). Revising the factor structure of the Simulator Sickness Questionnaire. *Annual Review of CyberTherapy and Telemedicine*, *5*, 128–137. Retrieved from http://vrphobia.com/Research/Publications/ARCTT2007.pdf#page=117

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

Chen, Y.-C., Dong, X., Hagstrom, J., & Stoffregen, T. A. (2011). Control of a virtual ambulation influences body movement and motion sickness. *BIO Web of Conferences*, *1*, 00016. https://doi.org/10.1051/bioconf/20110100016

Cobb, S. V. (1999). Measurement of postural stability before and after immersion in a virtual environment. *Applied Ergonomics*, *30*(1), 47–57. https://doi.org/10.1016/s0003-6870(98)00038-6

Davis, S., Nesbitt, K., & Nalivaiko, E. (2014). A Systematic Review of Cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment* (pp. 1–9). ACM. https://doi.org/10.1145/2677758.2677780

Davis, S., Nesbitt, K., & Nalivaiko, E. (2015). Comparing the onset of cybersickness using the oculus rift and two virtual roller coasters. In *Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015)* (Vol. 27, p. 30). Retrieved from http://crpit.com/confpapers/CRPITV167Davis.pdf

Dong, X., & Stoffregen, T. A. (2010). Postural Activity and Motion Sickness among Drivers and Passengers in a Console Video Game. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting*, *54*(18), 1340–1344. https://doi.org/10.1177/154193121005401808

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037//1082-989x.4.3.272

Fernandes, A. S., & Feiner, S. K. (2016). Combating VR sickness through subtle dynamic field-of-view modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 201–210). https://doi.org/10.1109/3DUI.2016.7460053

Golding, J. F. (1998). Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain Research Bulletin*, *47*(5), 507–516. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10052582

Graeber, D. A., & Stanney, K. M. (2002). Gender Differences in Visually Induced Motion Sickness. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting*, *46*(26), 2109–2113. https://doi.org/10.1177/154193120204602602

Graybiel, A., Wood, C. D., Miller, E. F., & Cramer, D. B. (1968). Diagnostic criteria for grading the severity of acute motion sickness. *Aerospace Medicine*, *39*(5), 453–455. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5648730

Hakkinen, J., Vuori, T., & Paakka, M. (2002). Postural stability and sickness symptoms after HMD use. In *IEEE International Conference on Systems, Man and Cybernetics* (Vol. vol.4, pp. 147–152). Yasmine Hammamet, Tunisia: IEEE. https://doi.org/10.1109/ICSMC.2002.1167964

HTC Vive Health and Safety Guide. (n.d.). Retrieved September 2016, from http://dl4.htc.com/web_materials/Safety_Guide/Vive/Vive_safety_and_regulatory_guide_ENG-FRC-ESM.pdf

HTC Vive Pre Installation Guide. (n.d.). Retrieved October 2016, from https://support.steampowered.com/kb_article.php?ref=2001-UXCM-4439

Iowa State Gender, Ethnicity & Residence Reports. (n.d.). Retrieved September 2016, from http://www.registrar.iastate.edu/enrollment/gender

Kellogg, R. S., Kennedy, R. S., & Graybiel, A. (1965). MOTION SICKNESS SYMPTOMATOLOGY OF LABYRINTHINE DEFECTIVE AND NORMAL SUBJECTS DURING ZERO GRAVITY MANEUVERS. *Aerospace Medicine*, *36*, 315–318. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14288892

Kelly, J. W., Riecke, B., Loomis, J. M., & Beall, A. C. (2008). Visual control of posture in real and virtual environments. *Perception & Psychophysics*, *70*(1), 158–165. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18306969

Kennedy, R. S., & Fowlkes, J. E. (1992). Simulator Sickness Is Polygenic and polysymptomatic: Implications for Research. *The International Journal of Aviation Psychology*, *2*(1), 23–38. https://doi.org/10.1207/s15327108ijap0201_2

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (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., Lilienthal, M. G., & Berbaum, K. S. (1989). Simulator sickness in US Navy flight simulators. *Aviation, Space, and Environmental Medicine*. Retrieved from http://europepmc.org/abstract/med/2923588

Kennedy, R. S., Tolhurst, G. C., & Graybiel, A. (1965). The effects of visual deprivation on adaptation to a rotating environment. NSAM-918. *Research Report. Naval School of Aviation Medicine*, 1–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5294754

Keshavarz, B., & Hecht, H. (2011). Validating an efficient method to quantify motion sickness. *Human Factors*, *53*(4), 415–426. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21901938

Keshavarz, B., Riecke, B. E., Hettinger, L. J., & Campos, J. L. (2015). Vection and visually induced motion sickness: how are they related? *Frontiers in Psychology*, *6*, 472. https://doi.org/10.3389/fpsyg.2015.00472

Kim, D. H., Parker, D. E., & Park, M. Y. (2004). *A New Procedure for Measuring Simulator Sickness–the RSSQ*.

Kiryu, T., Uchiyama, E., Jimbo, M., & Iijima, A. (2007). Time-Varying Factors Model with Different Time-Scales for Studying Cybersickness. In *Virtual Reality* (pp. 262–269). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73335-5_29

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

Liu, C.-L. (2009). A neuro-fuzzy warning system for combating cybersickness in the elderly caused by the virtual environment on a TFT-LCD. *Applied Ergonomics*, *40*(3), 316–324. https://doi.org/10.1016/j.apergo.2008.12.001

Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). Comparing the Fit of Item Response Theory and Factor Analysis Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(3), 333–356. https://doi.org/10.1080/10705511.2011.581993

McCauley, M. E., & Sharkey, T. J. (1992). Cybersickness: Perception of Self-Motion in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *1*(3), 311–318. https://doi.org/10.1162/pres.1992.1.3.311

Miller, E. F., 2nd, & Graybiel, A. (1970). A provocative test for grading susceptibility to motion sickness yielding a single numerical score. *Acta Oto-Laryngologica. Supplementum*, *274*, 1–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5280975

Miller, E. F., II, & Graybiel, A. (1974). Comparison of five levels of motion sickness severity as the basis for grading susceptibility. Retrieved from http://ntrs.nasa.gov/search.jsp?R=19740051130

Morean, M. E., Corbin, W. R., & Treat, T. A. (2013). The Subjective Effects of Alcohol Scale: development and psychometric evaluation of a novel assessment tool for measuring subjective response to alcohol. *Psychological Assessment*, *25*(3), 780–795. https://doi.org/10.1037/a0032542

Moss, J. D., & Muth, E. R. (2011). Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors*, *53*(3), 308–319. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21830515

Muhanna, M. A. (2015/7). Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions. *Journal of King Saud University - Computer and Information Sciences*, *27*(3), 344–361. https://doi.org/10.1016/j.jksuci.2014.03.023

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum Sample Size Recommendations for Conducting Factor Analyses. *International Journal of Testing*, *5*(2), 159–168. https://doi.org/10.1207/s15327574ijt0502_4

Muth, E. R., Stern, R. M., Thayer, J. F., & Koch, K. L. (1996). Assessment of the multiple dimensions of nausea: the Nausea Profile (NP). *Journal of Psychosomatic Research*, *40*(5), 511–520. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8803860

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. *Physiology & Behavior*, *151*, 583–590. https://doi.org/10.1016/j.physbeh.2015.08.043

Nichols, S., & Patel, H. (2002). Health and safety implications of virtual reality: a review of empirical evidence. *Applied Ergonomics*, *33*(3), 251–271. https://doi.org/10.1016/s0003-6870(02)00020-0

Oman, C. M. (1990). Motion sickness: a synthesis and evaluation of the sensory conflict theory. *Canadian Journal of Physiology and Pharmacology*, *68*(2), 294–303. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2178753

Plouzeau, J., Paillot, D., Chardonnet, J.-R., & Merienne, F. (2015). Effect of proprioceptive vibrations on simulator sickness during navigation task in virtual environment. *International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments*. Retrieved from http://sam.ensam.eu/handle/10985/10422

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

R. Philip Chalmers (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software, 48(6), 1-29. URL http://www.jstatsoft.org/v48/i06/.

Reason, J. T., & Brand, J. J. (1975). *Motion sickness*. Academic Press. Retrieved from http://psycnet.apa.org/psycinfo/1976-12574-000

Rebenitsch, L., & Owen, C. (2014). Individual variation in susceptibility to cybersickness. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (pp. 309–317). ACM. https://doi.org/10.1145/2642918.2647394

Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality*, *20*(2), 101–125. https://doi.org/10.1007/s10055-016-0285-9

Serge, S. R., & Moss, J. D. (2015). Simulator Sickness and the Oculus Rift: A First Look. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting*, *59*(1), 761–765. https://doi.org/10.1177/1541931215591236

Shupak, A., & Gordon, C. R. (2006). Motion sickness: advances in pathogenesis, prediction, prevention, and treatment. *Aviation, Space, and Environmental Medicine*, *77*(12), 1213–1223. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17183916

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Stanney, K. M., & Kennedy, R. S. (1998). Aftereffects from Virtual Environment Exposure: How Long do They Last? *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting*, *42*(21), 1476–1480. https://doi.org/10.1177/154193129804202103

Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997). Cybersickness is Not Simulator Sickness. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting*, *41*(2), 1138–1142. https://doi.org/10.1177/107118139704100292

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human Factors Issues in Virtual Environments: A Review of the Literature. *Presence* , *7*(4), 327–351. https://doi.org/10.1162/105474698565767

Stoffregen, T. A., & Riccio, G. E. (1991). An Ecological Critique of the Sensory Conflict Theory of Motion Sickness. *Ecological Psychology: A Publication of the International Society for Ecological Psychology*, *3*(3), 159–194. https://doi.org/10.1207/s15326969eco0303_1

Stout, C. S., & Cowings, R. S. (1993). Increasing accuracy in the assessment of motion sickness: A construct methodology (NASA TM-108797). *Ames Research Center, CA: NASA Space Life*.

Strauss, S. (1995). Cybersickness: The Side Effects of Virtual Reality. *Technology Review*, *98*(5), 14–16. Retrieved from http://dl.acm.org/citation.cfm?id=204577.204580

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

van Emmerik, M. L., de Vries, S. C., & Bos, J. E. (2011). Internal and external fields of view affect cybersickness. *Displays*, *32*(4), 169–174. https://doi.org/10.1016/j.displa.2010.11.003

VRHeads: Field of view face-off: Rift vs Vive vs Gear VR vs PSVR. (2016, July 25). Retrieved September 2016, from http://www.vrheads.com/field-view-faceoff-rift-vs-vive-vs-gear-vr-vs-psvr#vr

Warwick-Evans, L. A., Symons, N., Fitch, T., & Burrows, L. (1998). Evaluating sensory conflict and postural instability. Theories of motion sickness. *Brain Research Bulletin*, *47*(5), 465–469. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10052575

Wilson, J. R. (1997). Virtual environments and ergonomics: needs and opportunities. *Ergonomics*, *40*(10), 1057–1077. https://doi.org/10.1080/001401397187603

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

Young, S. D., Adelstein, B. D., & Ellis, S. R. (2006). Demand Characteristics of a Questionnaire Used to Assess Motion Sickness in a Virtual Environment. In *IEEE Virtual Reality Conference (VR 2006)* (pp. 97–102). https://doi.org/10.1109/VR.2006.44

APPENDIX A. TABLES

Table 1. VR game summary information.				
VR video game	**n**	**Game motion**	**Minutes of playtime**	**Would play again**
NoLimits Rollercoaster Simulation	18	discordant	8.9 (2.1)	9 (50%)
VR Karts	70	discordant	16.8 (5.1)	25 (36%)
Waltz of the Wizard	114	concordant	19.4 (2.1)	73 (64%)

Title of VR game. Number of participants that played that game. Game motion, either discordant or concordant. Mean minutes played (SD). Count (percentage) that reported they "definitely" or "probably" would play the game again.

Table 2. VR Game Evaluation rating item means, SDs, and response frequencies.							
	Mean	**SD**	**Strongly disagree**	**Somewhat disagree**	**Neither agree nor disagree**	**Somewhat agree**	**Strongly agree**
Entertaining	1.14	0.90	1	14	18	85	77
Exciting	0.82	0.99	4	20	31	92	48
Fun	1.06	0.96	3	15	21	85	71
Absorbing	0.53	1.17	17	14	55	66	43
Enjoyable	1.11	0.88	3	8	23	91	70
Involving	0.90	1.06	7	16	30	79	63
Challenging	-0.55	1.21	54	57	33	45	6
Frustrating	-0.92	1.12	78	56	31	27	3
Boring	-0.95	1.09	76	64	28	23	4
Difficult to play	-1.30	0.88	102	61	22	9	1
Strenuous	-1.16	1.08	106	37	31	19	2
Fatiguing	-1.18	1.09	107	42	23	20	3
Painful	-1.69	0.67	153	27	11	4	0
Disorienting	-0.42	1.32	59	38	34	53	11
Nauseating	-0.94	1.35	107	24	21	32	11
Dizzying	-0.61	1.40	79	36	16	52	12

Table 3. VR Game Evaluation rank order correlations between items (Kendall's tau).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Entertaining															
2 Exciting	.61														
3 Fun	.72	.66													
4 Absorbing	.31	.35	.33												
5 Enjoyable	.72	.59	.78	.40											
6 Involving	.40	.35	.43	.36	.43										
7 Challenging	.10	.15	.17	.09	.09	.29									
8 Frustrating	-.18	-.11	-.16	-.07	-.25	-.03	.30								
9 Boring	-.53	-.49	-.54	-.34	-.57	-.38	-.14	.11							
10 Difficult to play	-.10	-.04	-.02	.06	-.09	.04	.37	.29	.08						
11 Strenuous	-.16	-.12	-.18	.02	-.20	-.03	.23	.34	.12	.33					
12 Fatiguing	-.16	.00	-.14	.03	-.20	-.12	.16	.26	.13	.20	.44				
13 Painful	-.34	-.18	-.26	-.06	-.34	-.18	.08	.22	.19	.23	.42	.50			
14 Disorienting	-.17	-.10	-.17	.14	-.15	-.03	.14	.19	.07	.19	.37	.36	.30		
15 Nauseating	-.31	-.21	-.34	-.06	-.34	-.19	.00	.20	.22	.12	.34	.47	.53	.45	
16 Dizzying	-.29	-.18	-.29	-.07	-.30	-.18	.04	.23	.21	.16	.34	.49	.46	.53	.69

Table 4. VR Game Evaluation graded response models goodness of fit statistics.

Type	Factors	M2	RMSEA	TLI
Confirmatory	3	$M2(54) = 143.48$, $p = .00$	0.092 [0.074, 0.111]	0.917
Exploratory	4	$M2(15) = 21.09$, $p = .13$	0.046 [0, 0.087]	0.98

Model type, either confirmatory or exploratory graded response model. Number of factors fit by the model. M2, degrees of freedom, and p-value (Maydeu-Olivares and Joe, 2006). Root Mean Square Error of Approximation (RMSEA) with 95% confidence interval. Tucker-Lewis Index (TLI).

Table 5. VR Game Evaluation model standardized and rotated (oblimin) factor loadings.			

Confirmatory three-factor model.

	F1	F2	F3
Entertaining	.92	.	.
Exciting	.86	.	.
Fun	.95	.	.
Absorbing	.51	.	.
Enjoyable	.94	.	.
Involving	.59	.	.
Challenging	.	.65	.
Frustrating	.	.63	.
Boring	-.78	.	.
Difficult to play	.	.70	.
Strenuous	.	.	.60
Fatiguing	.	.	.70
Painful	.	.	.81
Disorienting	.	.	.71
Nauseating	.	.	.94
Dizzying	.	.	.92

Exploratory four-factor model.

	F1	F2	F3	F4
Entertaining	-.08	-.89	-.10	.12
Exciting	.10	-.89	-.14	.11
Fun	-.11	-.90	.01	-.05
Absorbing	.20	-.66	.20	-.17
Enjoyable	-.11	-.89	.16	-.05
Involving	-.08	-.61	-.10	-.16
Challenging	-.03	-.23	-.33	-.45
Frustrating	.14	.22	-.44	-.25
Boring	-.06	.81	-.06	.01
Difficult to play	.02	.04	.03	-.79
Strenuous	.50	.03	-.33	-.29
Fatiguing	.73	-.02	-.34	.05
Painful	.69	.20	-.07	-.19
Disorienting	.82	-.15	.13	-.08
Nauseating	.92	.09	.04	.05
Dizzying	.92	.05	.04	.03

Table 6. VR Game Evaluation exploratory four-factor model individual item fit statistics.

	S_X2	df.S_X2	p.S_X2
Entertaining	35.61990	31	.26
Exciting	43.09066	40	.34
Fun	39.56124	33	.20
Absorbing	58.77387	46	.10
Enjoyable	42.45831	33	.13
Involving	54.34681	41	.08
Challenging	43.36870	39	.29
Frustrating	48.00959	41	.21
Boring	82.01870	52	.00
Difficult to play	19.93462	26	.79
Strenuous	52.87241	26	.00
Fatiguing	49.41388	29	.01
Painful	25.39904	16	.06
Disorienting	40.64071	39	.40
Nauseating	43.15960	40	.34
Dizzying	29.92706	39	.85

Table 7. SSQ item means, SDs, and response frequencies.

	Mean	SD	None	Slight	Moderate	Severe
General discomfort	0.49	0.70	121	50	23	0
Fatigue	0.25	0.56	157	27	9	1
Headache	0.34	0.65	147	30	16	1
Eyestrain	0.60	0.75	106	63	22	3
Difficulty focusing	0.45	0.74	131	42	17	4
Increased salivation	0.18	0.45	164	25	5	0
Sweating	0.60	0.81	112	52	25	5
Nausea	0.46	0.83	139	28	19	8
Difficulty concentrating	0.26	0.54	152	33	9	0
Fullness of head	0.43	0.73	133	43	13	5
Blurred vision	0.38	0.74	145	31	12	6
Dizziness (eyes open)	0.31	0.70	154	24	11	5
Dizziness (eyes closed)	0.35	0.71	149	28	12	5
Vertigo	0.16	0.44	168	22	3	1
Stomach awareness	0.37	0.72	147	25	19	3
Burping	0.11	0.39	177	14	2	1

<table>
<thead>
<tr><th colspan="16">Table 8. SSQ rank order correlations between items (Kendall's tau).</th></tr>
<tr><th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th></tr>
</thead>
<tbody>
<tr><td>1 General discomfort</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2 Fatigue</td><td>.41</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>3 Headache</td><td>.42</td><td>.30</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>4 Eyestrain</td><td>.42</td><td>.34</td><td>.38</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>5 Difficulty focusing</td><td>.35</td><td>.26</td><td>.24</td><td>.42</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>6 Increased salivation</td><td>.21</td><td>.16</td><td>.12</td><td>.30</td><td>.29</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>7 Sweating</td><td>.30</td><td>.19</td><td>.11</td><td>.14</td><td>.08</td><td>.19</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>8 Nausea</td><td>.59</td><td>.41</td><td>.43</td><td>.37</td><td>.14</td><td>.17</td><td>.36</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>9 Difficulty concentrating</td><td>.41</td><td>.34</td><td>.42</td><td>.37</td><td>.58</td><td>.37</td><td>.12</td><td>.34</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>10 Fullness of head</td><td>.40</td><td>.34</td><td>.32</td><td>.40</td><td>.39</td><td>.32</td><td>.18</td><td>.27</td><td>.47</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>11 Blurred vision</td><td>.15</td><td>.19</td><td>.31</td><td>.45</td><td>.49</td><td>.24</td><td>.05</td><td>.19</td><td>.32</td><td>.31</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>12 Dizziness (eyes open)</td><td>.44</td><td>.40</td><td>.43</td><td>.50</td><td>.39</td><td>.27</td><td>.14</td><td>.47</td><td>.51</td><td>.36</td><td>.37</td><td></td><td></td><td></td><td></td></tr>
<tr><td>13 Dizziness (eyes closed)</td><td>.40</td><td>.38</td><td>.36</td><td>.43</td><td>.31</td><td>.26</td><td>.13</td><td>.47</td><td>.41</td><td>.32</td><td>.36</td><td>.82</td><td></td><td></td><td></td></tr>
<tr><td>14 Vertigo</td><td>.32</td><td>.21</td><td>.31</td><td>.36</td><td>.19</td><td>.13</td><td>.15</td><td>.32</td><td>.29</td><td>.25</td><td>.30</td><td>.43</td><td>.39</td><td></td><td></td></tr>
<tr><td>15 Stomach awareness</td><td>.47</td><td>.21</td><td>.26</td><td>.27</td><td>.22</td><td>.30</td><td>.35</td><td>.62</td><td>.34</td><td>.28</td><td>.20</td><td>.40</td><td>.40</td><td>.40</td><td></td></tr>
<tr><td>16 Burping</td><td>.17</td><td>.08</td><td>.21</td><td>.12</td><td>.09</td><td>.06</td><td>.16</td><td>.23</td><td>.20</td><td>.16</td><td>.18</td><td>.27</td><td>.29</td><td>.36</td><td>.28</td></tr>
</tbody>
</table>

<table>
<thead>
<tr><th colspan="8">Table 9. SSQ and CSQ graded response models goodness of fit statistics.</th></tr>
<tr><th>Measure</th><th>Model type</th><th>Items</th><th>Factors</th><th>M2</th><th>RMSEA</th><th>TLI</th></tr>
</thead>
<tbody>
<tr><td>SSQ</td><td>Confirmatory</td><td>16</td><td>3</td><td>$M2(70) = 447.4, p = .00$</td><td>0.167 [0.152, 0.182]</td><td>0.795</td></tr>
<tr><td>SSQ</td><td>Exploratory</td><td>16</td><td>4</td><td>$M2(33) = 44.82, p = .08$</td><td>0.043 [0, 0.072]</td><td>0.986</td></tr>
<tr><td>CSQ</td><td>Exploratory</td><td>7</td><td>2</td><td>$M2(1) = 1.62, p = .20$</td><td>0.057 [0, 0.209]</td><td>0.984</td></tr>
<tr><td>CSQ</td><td>Exploratory</td><td>9</td><td>3</td><td>$M2(3) = 2.47, p = .48$</td><td>0 [0, 0.113]</td><td>1.005</td></tr>
<tr><td>CSQ</td><td>Exploratory</td><td>9</td><td>2</td><td>$M2(10) = 4.68, p = .91$</td><td>0 [0, 0.03]</td><td>1.016</td></tr>
</tbody>
</table>

Model type, either confirmatory or exploratory graded response model. Number of factors fit by the model. M2, degrees of freedom, and p-value (Maydeu-Olivares and Joe, 2006). Root Mean Square Error of Approximation (RMSEA) with 95% confidence interval. Tucker-Lewis Index (TLI).

Table 10. SSQ confirmatory 16-item, three-factor model, structure based on Kennedy et al. (1993), model standardized and rotated (oblimin) factor loadings.

	F1 Nausea (N)	F2 Oculomotor strain (O)	F3 Disorientation (D)
General discomfort	.68	.42	.
Fatigue	.	.63	.
Headache	.	.68	.
Eyestrain	.	.83	.
Difficulty focusing	.	.87	-.07
Increased salivation	.41	.	.
Sweating	.63	.	.
Nausea	.87	.	.29
Difficulty concentrating	.24	.74	.
Fullness of head	.	.	.52
Blurred vision	.	.77	.04
Dizziness (eyes open)	.	.	1.00
Dizziness (eyes closed)	.	.	1.00
Vertigo	.	.	.66
Stomach awareness	.91	.	.
Burping	.61	.	.

Table 11. CSQ exploratory nine-item, two-factor model standardized and rotated (oblimin) factor loadings.

	F1 Dizziness	F2 Difficulty focusing
Headache	-.50	-.21
Eyestrain	-.33	-.58
Difficulty focusing	.07	-.89
Nausea	-.84	.14
Fullness of head	-.24	-.55
Blurred vision	-.01	-.81
Dizziness (eyes open	-.89	-.15
Dizziness (eyes closed)	-.99	.04
Vertigo	-.54	-.22

Item weights for scoring (factor loadings < .40 suppressed).

	F1 Dizziness	F2 Difficulty focusing
Headache	.50	.
Eyestrain	.	.58
Difficulty focusing	.	.89
Nausea	.84	.
Fullness of head	.	.55
Blurred vision	.	.81
Dizziness (eyes open	.89	.
Dizziness (eyes closed)	.99	.
Vertigo	.54	.

Table 12. CSQ exploratory nine-item, two-factor model individual item fit statistics.

	S_X2	df.S_X2	p.S_X2
Headache	12.871341	13	.46
Eyestrain	13.401220	10	.20
Difficulty focusing	17.267618	15	.30
Nausea	17.650560	15	.28
Fullness of head	8.127671	12	.78
Blurred vision	25.570938	14	.03
Dizziness (eyes open	1.911887	5	.86
Dizziness (eyes closed)	3.450385	6	.75
Vertigo	12.200753	8	.14

	1	2	3	4	5	6	7	8	9	10	11
1 MSSQ											
2 VG experience	.46										
3 VR experience	-.14	-.02									
4 Female	-.13	-.05	.03								
5 Discordant	.16	.02	-.33	-.10							
6 Normal vision	.17	.04	.00	.01	-.03						
7 Contacts	.04	.05	-.06	-.04	-.09	.09					
8 Eyeglasses	.05	-.04	-.06	.06	.06	-.04	-.56				
9 CSQ Dizziness	-.08	-.02	.13	-.01	.05	-.07	-.64	-.28			
10 CSQ Difficulty focusing	.36	.23	-.12	-.01	.08	.42	-.01	.02	-.01		
11 VRGE Nauseating	.27	.20	.01	-.06	.17	.22	-.06	.01	.06	.32	
12 NA	.30	.18	-.17	-.07	.12	.41	-.02	.05	-.03	.42	.29

Table 13. Correlations (Kendall's tau) between variables of interest for predicting cybersickness.

Table 14. Summary statistics for five logistic regression models with lowest AIC predicting CSQ Dizziness, CSQ Difficulty focusing, and VRGE Nauseating.

CSQ Dizziness

	Model	p	AIC	BIC	Hoslem	Hoslem.df	Hoslem.p
73	csq_dizzy~vis_norm+mssq+discordant	3	114.934	125.395	8.948	8	0.347
9	csq_dizzy~mssq+discordant	2	115.261	123.106	5.852	8	0.664
77	csq_dizzy~vis_norm+mssq+vg_hrs_wk+discordant	4	115.427	128.503	3.492	8	0.900
13	csq_dizzy~mssq+vg_hrs_wk+discordant	3	115.904	126.365	2.718	8	0.951
25	csq_dizzy~vis_glass+mssq+discordant	3	116.070	126.531	10.990	8	0.202

CSQ Difficulty focusing

	Model	p	AIC	BIC	Hoslem	Hoslem.df	Hoslem.p
21	csq_dif_foc~vis_glass+vg_hrs_wk+discordant	3	127.243	137.704	9.684	8	0.288
17	csq_dif_foc~vis_glass+discordant	2	127.456	135.301	1.465	8	0.993
149	csq_dif_foc~female+vis_glass+vg_hrs_wk+discordant	4	128.408	141.483	9.629	8	0.292
69	csq_dif_foc~vis_norm+vg_hrs_wk+discordant	3	128.753	139.214	9.670	8	0.289
53	csq_dif_foc~vis_cont+vis_glass+vg_hrs_wk+discordant	4	128.881	141.956	8.493	8	0.387

VRGE Nauseating

	Model	p	AIC	BIC	Hoslem	Hoslem.df	Hoslem.p
5	vrge_nause~vg_hrs_wk+discordant	2	80.247	88.093	4.865	8	0.772
7	vrge_nause~vg_hrs_wk+vr_experience+discordant	3	80.328	90.788	4.489	8	0.811
9	vrge_nause~mssq+discordant	2	80.346	88.191	6.107	8	0.635
15	vrge_nause~mssq+vg_hrs_wk+vr_experience+discordant	4	80.653	93.728	1.799	8	0.987
13	vrge_nause~mssq+vg_hrs_wk+discordant	3	80.697	91.158	6.415	8	0.601

Table 15. Logistic regression model summary for selected model of best fit predicting CSQ Dizziness.

```
Call:
glm(formula = csq_dizzy ~ vis_norm + mssq + discordant, family = "binomial",
    data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3767  -0.7610  -0.4555   0.8863   1.8251

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.45585    0.48371  -3.010  0.00261 **
vis_norm    -0.75761    0.50751  -1.493  0.13549
mssq         0.03849    0.01303   2.954  0.00313 **
discordant   2.15683    0.50155   4.300 1.71e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 140.01  on 100  degrees of freedom
Residual deviance: 106.93  on  97  degrees of freedom
AIC: 114.93

Number of Fisher Scoring iterations: 4
```

Table 16. Logistic regression model summary for selected model of best fit predicting CSQ Difficulty focusing.

```
Call:
glm(formula = csq_dif_foc ~ vis_glass + vg_hrs_wk + discordant,
    family = "binomial", data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1990  -1.0291   0.6739   0.9032   1.4852

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.69967    0.43935  -1.593  0.11127
vis_glass    1.38561    0.58555   2.366  0.01796 *
vg_hrs_wk    0.06808    0.04663   1.460  0.14429
discordant   1.22998    0.46591   2.640  0.00829 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 132.71  on 100  degrees of freedom
Residual deviance: 119.24  on  97  degrees of freedom
AIC: 127.24

Number of Fisher Scoring iterations: 4
```

Table 17. Logistic regression model summary for selected model of best fit predicting VRGE Nauseating.

```
Call:
glm(formula = vrge_nause ~ vg_hrs_wk + discordant, family = "binomial",
    data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1157  -0.7549  -0.2343  -0.1110   2.6427

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.34072    1.04285  -3.203  0.00136 **
vg_hrs_wk   -0.12034    0.07851  -1.533  0.12535
discordant   3.19386    1.05711   3.021  0.00252 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 97.664  on 100  degrees of freedom
Residual deviance: 74.247  on  98  degrees of freedom
AIC: 80.247

Number of Fisher Scoring iterations: 6
```

APPENDIX B. FIGURES

Figure 1. Currently available consumer VR HMD systems. Clockwise from top left: HTC Vive, Oculus Rift, Sony PlayStation VR, Samsung Gear VR.

Figure 2. Illustration of HTC Vive wireless handheld controllers.

Varimax Factors From SSQ Symptoms

SSQ Symptom	Varimax Factors			η^2
	N	O	D	
General discomfort	.65	.40	.18	.62
Fatigue	.15	.54	−.04	.32
Headache	.22	.53	.15	.35
Eyestrain	.00	.74	.17	.58
Difficulty focusing	−.01	.61	.43	.56
Increased salivation	.53	.21	.13	.34
Sweating	.31	.24	.08	.16
Nausea	.75	.08	.30	.66
Difficulty concentrating	.32	.39	.27	.33
Fullness of head	.12	.17	.37	.18
Blurred vision	.01	.36	.40	.29
Dizzy (eyes open)	.17	.07	.76	.60
Dizzy (eyes closed)	.17	.09	.65	.46
Vertigo	.18	.08	.37	.17
Stomach awareness	.64	.03	.21	.45
Burping	.41	.04	.22	.22
Eigenvalue	2.21	2.11	1.98	
Percent of variance	14	13	12	

Figure 3. SSQ items and varimax factor loadings, Table 2 from Kennedy et al. (1993).

Computation of SSQ Scores

SSQ Symptom[a]	Weight		
	N	O	D
General discomfort	1	1	
Fatigue		1	
Headache		1	
Eyestrain		1	
Difficulty focusing		1	1
Increased salivation	1		
Sweating	1		
Nausea	1		1
Difficulty concentrating	1	1	
Fullness of head			1
Blurred vision		1	1
Dizzy (eyes open)			1
Dizzy (eyes closed)			1
Vertigo			1
Stomach awareness	1		
Burping	1		
Total[b]	[1]	[2]	[3]

Score
$$N = [1] \times 9.54$$
$$O = [2] \times 7.58$$
$$D = [3] \times 13.92$$
$$TS^c = [1] + [2] + [3] \times 3.74$$

[a]Scored 0, 1, 2, 3. [b]Sum obtained by adding symptom scores. Omitted scores are zero. [c]Total Score.

Figure 4. SSQ items and scoring, Table 4 from Kennedy et al. (1993).

Figure 5. Demographic and health descriptors.

Figure 6. Motion Sickness Susceptibility Questionnaire (MSSQ) scores.

Figure 7. Video game play per typical week (in hours) and prior virtual reality experience.

Figure 8. Example of EDA time-series with lines of best fit per epoch.

Figure 9. Empirical distributions of slope values for linear models fit to EDA time-series, per epoch.

Figure 10. Interest in playing the virtual reality game again.

Figure 11. VR Game Evaluation items empirical distributions.

Complete cases, n = 195

Figure 12. VR Game Evaluation scree plot for parallel analysis of polychoric matrix.

VRGE Nauseating item characteristic curves

VRGE Nauseating test information

VRGE Fun item characteristic curves

VRGE Fun test information

VRGE Difficult item characteristic curves

VRGE Difficult test information

Figure 13. VR Game Evaluation item characteristic curves and test information plots for each factor.

Color codes equate to: P1="Strongly disagree", P2="Somewhat disagree", P3="Neither agree nor disagree", P4="Somewhat agree", P5="Strongly agree".

Figure 14. VR Game Evaluation, empirical distributions of ratings for the four factors, based on the exploratory four-factor model.

Figure 15. SSQ items empirical distributions.

Complete cases, n = 194

CSQ Dizziness item characteristic curves

CSQ Dizziness test information

CSQ Difficulty focusing item characteristic curves

CSQ Difficulty focusing test information

Figure 16. CSQ item characteristic curves and test information plots for each factor.

Color codes equate to: P1="None", P2="Slight", P3="Moderate/Severe" (amalgamated).

Figure 17. SSQ factors empirical distributions of scores.

Figure 18. Scatterplots of willingness to play again and VRGE "Fun", by each CSQ factor, "Dizziness" and "Difficulty focusing".

Figure 19. Scatterplot of each CSQ factor, "Dizziness" and "Difficulty focusing" (scores based on standardized loadings from Table 11).

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Date: 12/22/2016

To: William Stone
W112 Lagomarcino

CC: Dr. Douglas A Gentile
W112 Lagomarcino Hall

From: Office for Responsible Research

Title: Attitudes about video games

IRB ID: 16-543

Approval Date:	12/21/2016	**Date for Continuing Review:**	12/20/2018
Submission Type:	New	**Review Type:**	Expedited

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**

- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required.

- **Obtain IRB approval prior to implementing <u>any</u> changes** to the study by submitting a Modification Form for Non-Exempt Research or Amendment for Personnel Changes form, as necessary.

- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.

- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.

- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Please be aware that IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. **Approval from other entities may also be needed.** For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **IRB approval in no way implies or guarantees that permission from these other entities will be granted.**

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 202 Kingland, to officially close the project.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.