STARS

University of Central Florida
STARS

Electronic Theses and Dissertations, 2004-2019

2006

Identification Of System Design Features That Affect Sickness In Virtual Environments

Julie Drexler
University of Central Florida

 Part of the Engineering Commons
Find similar works at: https://stars.library.ucf.edu/etd
University of Central Florida Libraries http://library.ucf.edu

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Drexler, Julie, "Identification Of System Design Features That Affect Sickness In Virtual Environments" (2006). *Electronic Theses and Dissertations, 2004-2019*. 797.
https://stars.library.ucf.edu/etd/797

IDENTIFICATION OF SYSTEM DESIGN FEATURES THAT AFFECT SICKNESS IN VIRTUAL ENVIRONMENTS

by

JULIE MARIE DREXLER
B.A. University of Central Florida, 1993
M.S. University of Central Florida, 2000

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2006

Major Professor: Dr. Linda C. Malone

© 2006 Julie M. Drexler

ABSTRACT

The terms “simulator” and “VR” are typically used to refer to specific types of virtual environments (VEs) which differ in the technology used to display the simulated environment. While simulators and VR devices may offer advantages such as low cost training, numerous studies on the effects to humans of exposure to different VEs indicate that motion sickness-like symptoms are often produced during or after exposure to the simulated environment. These deleterious side effects have the potential to limit the utilization of VE systems if they jeopardize the health and/or safety of the user and create liability issues for the manufacturer.

The most widely used method for assessing the adverse symptoms of VE exposure is the Simulator Sickness Questionnaire (SSQ). The method of scoring the symptoms reported by VE users permits the different sickness symptoms to be clustered into three general types of effects or subscales and the distribution or pattern of the three SSQ subscales provides a profile for a given VE device. In the current research, several different statistical analyses were conducted on the SSQ data obtained from 21 different simulator studies and 16 different VR studies in order to identify an underlying symptom structure (i.e., SSQ profile) or severity difference for various types of VE systems.

The results of the research showed statistically significant differences in the SSQ profiles and the overall severity of sickness between simulator and VR systems, which provide evidence that simulator sickness and VR sickness represent distinct forms of motion sickness. Analyses on three types of simulators (i.e., Fixed- and Rotary-Wing flight simulators and Driving

simulators) also found significant differences in the sickness profiles as well as the overall severity of sickness within different types of simulator systems. Analyses on three types of VR systems (i.e., HMD, BOOM, and CAVE) revealed that BOOM and CAVE systems have similar sickness profiles, which are different than the HMD system profile. Moreover, the results showed that the overall severity of sickness was greater in HMD systems than in BOOM and CAVE systems.

Recommendations for future research included additional psychophysical studies to evaluate the relationship between various engineering characteristics of VE systems and the specific types of sickness symptoms that are produced from exposure to them.

This dissertation is dedicated in loving memory of my dear friend, Helen L. Foard, who always believed in me. Although she is no longer with me in body, she will always be with me in spirit!

ACKNOWLEDGMENTS

During the years that I have spent in graduate school, I have come to believe that completing a Ph.D. program is based on the universal rule that, "what does not kill you, makes you stronger". However, during this time I also realized that it is not just acquiring knowledge that has helped me to grow, but also the people in my life that I have learned from and which have considerably enriched my life. I am eternally grateful to those who had the patience and wisdom to be my greatest teachers.

First, I would like to express my sincerest thanks to the members of my committee for their support, suggestions, and for taking time from their busy schedules to contribute to the success of this dissertation. Extreme appreciation and gratitude is given to my dissertation advisor and committee chair, Dr. Linda Malone, who guided me through every step of this project. She is an extraordinary women, gifted teacher, wonderful mentor, and caring friend and it has been a tremendous pleasure to work with her. Her kind heart, humor, and willingness to give me the freedom to find my own way, but always being available to help me when I needed it most will always be appreciated. I am particularly thankful for her statistical expertise, thoughtful feedback, careful editing of my work, as well as her endless support and personal encouragement throughout the dissertation process.

My sincere gratitude is also extended to Dr. Robert Kennedy for introducing me to, and cultivating my interest in, the field of human factors, teaching me how to be a scientist, and providing me with professional development opportunities for which most graduate students

could only dream. As a committee member, he not only provided the impetus for this research topic, but also provided access to the data that made this research project possible. I am very appreciative of his knowledge and guidance on the "sickness" literature, critical editing, insightful feedback, and scientific expertise. Deepest appreciation is given to Dr. Pamela McCauley-Bell, for always demonstrating her confidence in my abilities and supporting my research interests. Thank you for your encouragement, support, enthusiasm, and for your participation as a committee member. I am also very grateful to Dr. Mustapha Mouloua for his advice, encouragement, and helpful suggestions as a member of the committee. His experience, knowledge, and valuable comments served to enhance the quality of this project. Last, but not least, I would like to acknowledge Dr. Mansooreh Mollaghasemi for her time, encouraging attitude, and participation on my committee. I consider myself fortunate to have had the opportunity to work with, and learn from, all of these highly intelligent, caring, and motivated individuals.

This degree also would not have been possible without my wonderful family. I am incredibly thankful for my wonderful husband, Dean, who through more years than I can count has shared the joys, frustrations, anxiety attacks, tears, sleepless nights, spoiled evenings, and general lack of a real life while I have been in school. Thank you for all of your love, friendship, and laughter, particularly when the going got tough. I cannot tell you how much I appreciated your constant encouragement and understanding, especially during the times when I had to spend more time at the computer than with you. Your patience throughout my quest to "learn" symbolizes a true love for which I am forever grateful. Enormous appreciation to my parents, Janet and Fred, for teaching me that with the love, support, and strength of family, *anything* is attainable and for always being there for me through each and every joy and disappointment of

my life. Thank you for providing the opportunity and encouragement to let me choose my own way, allowing me to reach as high as I could, and instilling in me the confidence that I needed to be successful in life. Your pride and unwavering faith in my ability to meet each of my goals was a strong incentive for my eventual success. I am incredibly proud and fortunate to have such loving and supportive parents and I cannot thank you enough for everything that you have done for me. My appreciation and gratitude is also extended to my brother, Eric, for his devotion, humor, an always-listening ear, and constant encouragement during my many years in school. Sincere appreciation and heartfelt thanks also go to my grandmother, Lillian, and my stepmother, Joyce, for all of the caring, support, and encouragement that you have provided over the years. Words are inadequate to express my appreciation for the unconditional love and emotional support that each of you have provided to me throughout the years, especially during the times that I needed it most. You all have my sincere gratitude for your tireless efforts in supporting me during my graduate career and for enduring my absence while I completed the program requirements, particularly during the writing of this dissertation.

I am also very grateful to my friends who supported and encouraged me and who were never farther away than e-mail or the telephone. I have been blessed to have learned about unconditional friendship from Dr. Deborah Carstens. I am indebted to you for always unselfishly giving me your wisdom; offering a shoulder for me to lean on; providing comforting words, laughter, and unwavering support when I needed it most; and for *constantly* reminding me that I would finish this degree. Thank you for your caring, encouragement, laughter, and friendship throughout the years. I would also like to acknowledge my friends that I also had the pleasure of working with at RSK Assessments, Inc. My warmest appreciation goes to Daniel Compton for his continuous kindness, friendship, and always being available to answer any of

my hardware and software-related questions. Without his help and computer expertise, I would have been at a technological loss. An enormous thank you also goes to my friend Kimberly Sprouse who was always available to provide emotional support, encouragement, and laughter when I needed it most. I'm also not sure what I would have done if she had not kept me well supplied with Chipotle burritos and chips during the last few weeks before my dissertation was due. Kim also deserves special recognition for her invaluable assistance with sending out the data request letters, formatting several of the tables, and always finding "a few more articles" that I needed. I will never be able to thank you enough for being a true friend that was always willing to help me with whatever I needed to complete my dissertation. I would also like to acknowledge Dr. Sherrie Jones for her advise, encouragement, and friendship as well as loaning me her "sickness" library and assistance with finding information on the simulator equipment features.

Finally, I would like to acknowledge all of the people who graciously contributed their data to this project. You have my sincerest gratitude for your willingness to provide the assistance that made dissertation possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

BOOM	Binocular Omni-Oriented Monitor
CAVE	Cave (Computer) Automated Virtual Environment
CGI	Computer-Generated Imagery
CRT	Cathode Ray Tube
FOV	Field-of-View
HMD	Helmet (Head)-Mounted Display
HMMWV	High Mobility, Multipurpose, Wheeled Vehicle
IPD	Interpupillary Distance
IRB	Institutional Review Board
LCD	Liquid Crystal Display
LD	Labyrinthine Defective
MSI	Motion Sickness Incidence
MSQ	Motion Sickness Questionnaire
NASA	National Aeronautics and Space Administration
NATOPS	Naval Air Training and Operating Procedures Standardization
PAT	Pre-Flight Adaptation Trainer
SSQ	Simulator Sickness Questionnaire
VE	Virtual Environment
VR	Virtual Reality

CHAPTER ONE: INTRODUCTION

Virtual environment (VE) systems allow an individual to experience and interact with a simulated world. Through the use of computer-generated images, VE technology can generate vicarious and perceptually realistic images of a dynamic simulated environment. As the VE user navigates through the environment and interacts with virtual objects, this technology permits the user to experience a feeling of movement through the artificial world while remaining physically stationary.

There is no standard or generally accepted definition for virtual environments (VEs), simulators, or virtual reality (VR) devices (Blade & Padgett, 2002). Multiple uses of the terms VE and VR can be found in the literature and many authors often use these terms interchangeably. Accordingly, some of the terminology used throughout this dissertation must first be defined to prevent confusion or ambiguity in the terms. A "VE" is broadly defined as a device that presents users with a simulated environment where the user can interact with computer-generated images. The terms "simulator" and "VR" are used to refer to specific types of VEs which differ in the technology used to display the simulated environment. A simulator is a device that, in general, presents two-dimensional computer-generated scenes on a fixed-screen display such as a cathode ray tube (CRT), dome, or wrap-around projection screen. In contrast, a VR system employs a visually-coupled device, such as a helmet-mounted display (HMD) or stereographic glasses, that is worn by the user which typically present three-dimensional images.

Statement of the Problem

Although VE technology is rapidly progressing (e.g., computer speed, image generation, etc.), the deleterious side effects associated with VE exposure are still a major problem facing the VE industry. State-of-the-art and compellingly realistic VE systems currently exist, but very few of these devices do not induce sickness. For example, Kennedy and Stanney (1997) reported that 30-40% of flight simulator users reported being asymptomatic and only 5-10% of VR users did not report symptoms. The pervasiveness of sickness and the corresponding health and safety consequences related to the adverse effects of exposure to these devices clearly limits the use of existing VE applications. Furthermore, if the problem is not adequately addressed in the future, proposed VE applications will be adversely impacted and the development of future VE systems may be compromised. Thus, a critical and unresolved human factors issue associated with VE systems is the prevalence of the adverse effects that occur during and/or after exposure to a simulated environment.

An essential factor to understanding and ultimately solving the problem of VE sickness lies in the design of the systems. Cobb, Nichols, Ramsey, and Wilson (1999) conducted a series of experiments with different VR system configurations (HMDs, computer processor speeds, tracker delays, etc.) and found differences in sickness symptomatology between experimental conditions. The authors concluded that research efforts should be directed toward identifying the equipment configurations that provoke sickness side effects. However, while research is available on various causes of the adverse effects of VE exposure, there is limited knowledge concerning the effects of system design variables on sickness, even though it has been implicated as a major factor in VE sickness. Specifically, there is no guiding theory as to which VE system features affect different types of sickness symptoms. Thus, the potential exists to develop a

theory that could be used to identify system design features which will provoke different types of symptomatology and provide design strategies that could be employed to control the adverse effects of VE exposure.

Kennedy, Lanham, Drexler, Massey, and Lilienthal (1997) suggested that the first technical step toward improving VE systems so that they do not induce sickness is to quantify, as accurately as possible, the problem(s) that are experienced by the people who use them. The most widely used method for assessing and quantifying the adverse effects of VE exposure is the Simulator Sickness Questionnaire (SSQ). The method of scoring the symptoms reported by VE users permits the different sickness symptoms to be clustered into three general types of effects or subscales (Kennedy & Lilienthal, 1994). The intent of the symptom clustering was to provide diagnostic information which could be used to identify system design characteristics that influence the symptoms experienced by VE system users (Kennedy, Lane, Berbaum, & Lilienthal, 1993). Specifically, differences in the distribution or pattern of the three SSQ subscales (i.e., profile differences) may indicate the nature of the sickness engendered by a given VE device (Kennedy, Drexler, Stanney, & Harm, 1997).

The objective of the current research was to identify an underlying symptom structure (i.e., SSQ profile) for different types of VE systems and then determine whether there were quantitative differences in the patterns of symptoms over diverse systems. Kennedy, Drexler, Stanney, and Harm (1997) indicated that similarities in SSQ symptom profiles from two different environments would suggest a common cause, even if the similar profile occurred in a different VE with different visual display systems or other design characteristics. Likewise, the authors suggested that SSQ profile differences (e.g., excessive visual disturbance) may signal differences in specific equipment design features that differentially affect the severity and types of

symptoms reported. Thus, one of the goals of the research was to determine the form of the relationship between different engineering features and sickness symptoms and evaluate the generalizability of the relationships between sickness profiles and system features over different VE devices.

Additionally, the terms cybersickness or virtual reality sickness are commonly used in the VE sickness literature to refer to the adverse effects produced by VR devices in order to distinguish the symptoms from those produced by simulators. Two different terms, however, imply two distinct forms of motion sickness. A fundamental question that has not been addressed is whether simulator sickness and cybersickness produce sufficiently different types of symptoms to justify the use of two separate terms. Therefore, a second objective of the research was to determine whether the sickness produced by exposure to simulators and VR devices were quantitatively different by comparing the SSQ profiles obtained from different simulators and virtual reality devices and thereby provide evidence as to whether they represent distinct motion sickness constructs.

This research was necessitated by the need for non-system specific information on the design features that are best suited to minimize particular types of symptoms related to VE exposure. An understanding of the differential effects of various equipment features on sickness outcomes therefore, will facilitate effective management of VE-induced sickness (i.e., minimize side effects) in several different areas. The results of the research are intended to assist engineers, system designers, manufacturers, as well as owners and users of VE systems to reduce the potential health and safety consequences associated with the side effects of exposure to VE systems.

CHAPTER TWO: LITERATURE REVIEW

Applications of Virtual Environment Technology

Due to the maturity and flexibility of VE technology, a wide range of VE applications are currently available in military, medical, educational, commercial, and industrial settings. Because the visual images are compellingly realistic and users can be exposed to scenarios that would be dangerous or impractical in the real environment, VE systems can provide a safe and highly cost effective alternative to real-world training. Moreover, there is considerable evidence to suggest that VE technology can enhance task performance in a training environment (Pepper, Smith, & Cole, 1981; Witmer, Bailey, & Knerr, 1996; Magee, 1995; Regian, Shebilske, & Monk, 1993; Kenyon & Afenya, 1995). The U.S. military has exploited VE technology for procedural training such as maintenance, submarine ship handling tasks, and weapon system operation (Munro, Breaux, Patrey, & Sheldon, 2002; Stone, 2002). VE technology has also been applied to military operational planning and mission rehearsal, tactical skill training, combat vehicle system operation (e.g., aircraft, tanks, ships), as well as training for non-combat missions such as crowd control, humanitarian assistance, and hazardous material situations (Knerr, Breaux, Goldberg, & Thurman, 2002). In addition to training individuals, the military has employed VE technology to train team skills. Teams consisting of two or more people, which may include both human and simulated (virtual) members, can be simultaneously trained in simulated environments, even if team members are geographically distributed (Salas, Oser,

Cannon-Bowers, & Daskarolis-Kring, 2002). NASA has also increased the use of VE technology for astronaut training, mission planning and rehearsal, International Space Station operations (Covault, 1998) and training extravehicular activities such as repair of the Hubble telescope (Stone, 2002).

In the medical field, VE technology has been adopted to train surgical techniques and medical skills such as intravenous (IV) needle insertion, diagnosis (e.g., virtual endoscopy), and preoperative planning and rehearsal of complicated surgical procedures (Satava & Jones, 2002, 2003). Clinical applications, particularly in the areas of neuropsychology and psychiatry, have also been developed. Applications in clinical neuropsychology typically focus on the assessment and rehabilitation of cognitive and functional impairments due to neurological disorders (e.g., Alzheimer's disease, dementia), learning or developmental disabilities, and traumatic brain injury (Rizzo, Buckwalter, & van der Zaag, 2002; Riva, Wiederhold, & Molinari, 2000). An example of a VE application for developmental and learning disabled individuals is the Virtual Life Skills project which provides a virtual world where individuals can learn and practice functional activities of daily life that are necessary for independent living (Cobb, Neale, Crosier, & Wilson, 2002; Cobb, Neale, & Reynolds, 1998). The virtual world offers a safe and controlled environment for users to learn skills such as safely crossing streets, food preparation, shopping at a grocery store, and how to use public transportation (Brown, Kerr, & Bayon, 1998; Neale, Brown, Cobb, & Wilson, 1999). In the psychiatric field, VE technology has been used to treat psychological disorders such as posttraumatic stress, obsessive-compulsive behavior, attention deficit disorder as well as for the treatment of specific phobias (North, North, & Coble, 2002). By exposing individuals to realistic representations of a particular anxiety producing stimuli, virtual therapy has been successfully used to systematically desensitize individuals to phobias

such as claustrophobia (fear of confined spaces), agoraphobia (fear of open or public places), acrophobia (fear of flying; Hodges et al., 1995), and fear of heights (Riva, Botella, Légeron, & Optale, 2004; North, North, & Coble, 2002).

The automotive industry has used VE technology as a flexible tool for vehicle design, human factors design and evaluation of automobile interior design, and developments of new automotive systems such as ABS, on-board aid systems, and adaptive cruise control (Bernasch & Haenel, 1995; Servignat, Flores, Kemeny, & Vernet, 1995). Automobile simulators have also been an important tool in driver training, assessment, and rehabilitation, particularly in the elderly (Moldenhauer, 1995; Triggs & Fronsko, 1995). Additionally, the use of automobile simulators has enabled researchers to study various aspects of driving including physiological behaviors (e.g., heart rate; Malaterre, 1995), and driver behaviors (e.g., steering-wheel operation; Boulanger & Chevennement, 1995). Immersive VE systems are also being developed for industrial applications such as facility layout and design, process planning, design of optimal workstation layout and work methods, and operator and maintenance training (Shewchuk, Chung, & Williges, 2002; Stone, 2002; Wilson, 1999). Similarly, commercial companies such as Boeing have used VE technology for simulation-based design efforts such as prototyping and evaluation of interior cabin and cockpit designs (Stone, 2002).

The entertainment industry has also exploited VE technology to produce interactive computer games and dynamic rides, which are available to a wide range of consumers. For instance, an indoor VE-based theme park called Disney Quest was recently opened in Orlando, FL. The VE technology used to create the various immersive and interactive virtual worlds within the park range from motion-based simulators to HMDs combined with a motion-based seat (Badiqué et al., 2002). For example, one of the rides enables users to design their own roller

coaster and then "ride" their design in a motion-based simulator. Other simulator-type rides, which are typically found in amusement arcades or theme parks, present video images that place the user inside of a vehicle (automobile, plane, roller coaster, etc.). These systems often use hydraulic systems, which are synchronized with the video image, to move the simulator platform in order to provide more realism in the simulation (Badiqué et al., 2002).

Recognizing the benefits of VE technology (e.g., interactivity and immersion), academic settings ranging from elementary school to college level have also developed VE applications to teach students a broad range of subjects. Specific educational applications have included: cell biology, architectural design, space science, spatial problem solving (Youngblut, 1998), electrostatic forces and fields, biological resource cycles (Moshell & Hughes, 2002), chemical engineering (Bell & Fogler, 1998), and other difficult science concepts such as Newton's law (Salzman, Dede, & Loftin, 1995). Finally, in the area of information visualization, VE technology is considered to be a valuable tool because it allows the exploration and interaction with large multidimensional, numeric datasets and facilitates the identification of meaningful relationships within a complex dataset, particularly time-varying data (Bryson, 2002). Relatedly, VE technology has been applied in battlefield visualization which allows military personnel to efficiently and effectively visualize a rapidly changing battlefield in order to plan and direct various battlefield operations (Hix et al., 1999).

Effects of VE Exposure

While simulators and VR devices may offer advantages such as low cost training, numerous studies on the effects to humans of exposure to different virtual environments indicate that human exposure to devices which present rearranged or altered perceptual worlds often

produce motion sickness-like symptoms during or after exposure to the simulated environment (Kennedy, Drexler, Compton, Stanney, Lanham, & Harm, 2003). The symptoms that typically occur as a result of exposure to VEs include disorientation, nausea, dizziness, sweating, drowsiness, eyestrain, headache, loss of postural stability, and vomiting, although infrequent, and the severity of the side effects can range from mild discomfort to debilitating illness (Drexler, Kennedy, & Compton, 2004; Kennedy, Fowlkes, & Lilienthal, 1993).

The motion-sickness like symptoms associated with exposure to flight simulators, known as simulator sickness, have been a problem for over forty years (Kennedy, Drexler, & Compton, 1997). In the first published report of simulator sickness, Miller and Goodson (1960) indicated that 78% of the flight students and instructors experienced some degree of sickness as a result of exposure to a military helicopter simulator. Since then, similar side effects have been associated with exposure to other types of flight simulators including fighter, transport, patrol, and attack aircraft (Crowley, 1987; Department of the Navy, 2004; Kennedy, Lilienthal, Berbaum, Baltzley, & McCauley, 1989; McCauley, 1984; Ungs, 1988) and vehicle simulators such as automobiles and tanks (Casali & Wierwille, 1980; Curry, Artz, Cathey, Grant, & Greenberg, 2002; Lampton, Kraemer, Kolasinski, & Knerr, 1995; Lerman et al., 1993). While these effects have been well documented in simulators, motion sickness-like symptoms have been increasingly reported by a significant proportion of VR users, particularly those devices which employ HMDs (Hettinger, 2002; Howarth & Costello, 1996; Kennedy, Jones, Lilienthal, & Harm, 1994; Moshell, Blau, Knerr, Lampton, & Bliss, 1993; Pausch, Crea, & Conway, 1992; Regan & Price, 1994). In order to distinguish the symptoms that occur from exposure to a VR device from simulator-induced symptoms, some authors have referred to the side effects of VR devices as virtual reality sickness or cybersickness (McCauley & Sharkey, 1992).

Significance of Side Effects from VE Exposure

The deleterious side effects of VE exposure have the potential to limit the utilization of VE systems, particularly as a training device, if they jeopardize the health and/or safety of the user and create liability issues for the manufacturer. If humans are unable to effectively function in the VE, training objectives may be compromised or could result in a negative transfer of training effect which could affect subsequent performance on the real-world task (Canaras, Gentner, & Schopper, 1995; Lathan, Tracey, Sebrechts, Clawson, & Higgins, 2002). McCauley (1984) pointed out that symptoms experienced while in a simulator could distract users and/or decrease their motivation during a training exercise and ultimately compromise the effectiveness of the training protocol (cf., Hettinger, Berbaum, Kennedy, Dunlap, & Nolan, 1990; Kennedy, Hettinger, & Lilienthal, 1990). Users that experience symptoms during a simulation may also learn new behaviors (i.e., coping mechanisms) such as minimizing head movements, using only the instruments (i.e., not looking at the visual displays), or avoiding aggressive maneuvers in order to avoid or reduce sickness symptoms (Baltzley, Kennedy, Berbaum, Lilienthal, & Gower, 1989; Hettinger, Berbaum, Kennedy, Dunlap, & Nolan, 1990; Kennedy, Hettinger, & Lilienthal, 1990; Kennedy, Lilienthal, Berbaum, Baltzley, & McCauley, 1989). However, while these behaviors may be appropriate for the simulated task, they may not necessarily be appropriate for performing the corresponding tasks in the real world (Lathan et al., 2002; Pausch, Crea, & Conway, 1992). McCauley (1984) and Pausch, Crea, and Conway (1992) also suggested that any negative transfer of training to the real-world device could cause the user to lose confidence in the training they receive from the simulator, resulting in decreased simulator usage. Similarly, once a user experiences simulator sickness, he/she may be reluctant to return to the simulator for subsequent training or, alternatively, could disengage some of the simulator features (e.g., the

motion base) to reduce the potential sickness (Crowley, 1987; McCauley, 1984). Moreover, if the sickness problem is too severe and cannot be remedied, the device could be discarded, like the helicopter simulator reviewed in Miller and Goodson (1960). For the company that owns the VE system, both of these situations have economic implications associated with the purchase of equipment, either specific components or the entire system, that cannot be used.

The utilization of VE systems for research applications could also be compromised by the presence of these symptoms. Individuals that are experiencing side effects may be unwilling or unable to remain in the environment. Consequently, a proportion of those exposed may prematurely cease their interaction with the VE device and withdraw from the study prior to its completion. For example, the withdrawal rate for a series of 13 VR studies, conducted by the U.S. Army Research Institute, ranged from 0-25% (Knerr et al., 2002) while VR studies conducted by Stanney and collegues reported a 12-19% early withdrawal rate (Stanney, Lanham, Kennedy, & Breaux, 1999; Stanney, Kingdon, & Kennedy, 2002). Moreover, Stanney, Kingdon, and Kennedy (2002) found almost a 50% withdrawal rate for participants in a 60-minute exposure group. Early participant withdrawal can result in higher research costs due to the need to test additional participants, delays in data collection and analyses, and if the project is sponsored by an outside funding agency, the potential for contract default if the high attrition rate affects the project completion date. Additionally, if the sickness problem is relatively severe, participant recruitment can be hindered once other potential volunteers hear about people getting sick during the study. There is also concern that the sickness resulting from VE exposure may compromise the continued development and use of VE technology (Stanney, Mourant, & Kennedy, 1998).

The side effects of exposure to VE systems also have the potential to jeopardize the health and/or safety of users. One such threat is the persistence of symptoms (i.e., aftereffects) for a prolonged period of time following termination of exposure to the system. Baltzley, Kennedy, Berbaum, Lilienthal, and Gower (1989) investigated the time course of recovery from simulator sickness and found 75% of the pilots that experienced symptoms indicated the symptoms dissipated within one hour after simulator exposure. Of greater concern to user safety, however, was the authors' findings which indicated that 13% of all military pilots exposed to different flight simulators reported aftereffects that persisted for more than four hours after exposure to the device and 8% of the pilots experienced symptoms for six or more hours. Likewise, Stanney and Kennedy (1998) reported persistent aftereffects from exposure to a VR system. In their study, the authors found significant levels of sickness symptomatology were still being reported one hour after participants ceased exposure to the device. Specifically, their results indicated that compared to pre-exposure levels, disorientation-type symptoms (e.g., dizziness) were 95 times higher, gastrointestinal related symptoms (e.g., nausea) were ten times higher, and visual disturbances (e.g., eyestrain) were seven times higher. Unfortunately, the study was not designed to evaluate the time course of symptom recovery beyond the one hour post-exposure period.

There have also been reports of extreme cases of prolonged VE aftereffects. In one case, Viirre and Ellisman (2003) reported that after a researcher used a desktop VE for ten minutes, the user only experienced postural instability for a few minutes immediately after exposure. But, several hours later, there was an onset of vertigo and nausea which persisted for *four* days. In an even more extreme case, a man was exposed to four different immersive VE rides over a period of 45 minutes (Kennedy, Stanney, & Fernandez, 1999). Due to side effects, including nausea,

vertigo, dizziness, drowsiness, and headache, he immediately left the VE facility and went home to bed. The incidence and extreme severity of the symptoms, particularly the vertigo, persisted for several months and full recovery did not occur until seven months after the initial exposure to the VE rides! In both of these extreme cases, the authors reported that no organic cause for the prolonged symptoms were found; physical examination of the inner ear and neurological functioning was normal. The fact that in both cases the symptoms ultimately subsided also implied a functional disorder rather than an organic cause.

Additional threats to user safety occur when the side effects of VE exposure appear after the user has left the VE facility. One potential safety hazard is delayed effects; a user is symptom-free during or immediately following exposure to a simulated environment, but symptom onset occurs during some period of time subsequent to stimulus exposure (Baltzley et al., 1989). For example, Miller and Goodson (1960) reported that while most of the individuals exposed to a helicopter simulator experienced sickness symptoms during the exposure, some users did not experience any symptoms until several hours after leaving the simulator. Of particular concern for users' safety was the authors' report of a flight instructor who was forced to stop his car and walk around in order to reduce the disorientation he was experiencing as a delayed effect of his earlier exposure to the simulator. Flashbacks also present a threat to user safety. Flashbacks occur when symptoms cease once exposure to a provocative stimulus is terminated, but symptom onset suddenly reoccurs later (Baltzley et al., 1989). McCauley (1984) cited a study by Kellogg et al. (1980) where pilots reported visual flashbacks that occurred eight to ten hours after exposure to a fixed-base flight simulator. Similarly, Stanney and Kennedy (1998) found that approximately 31% of the participants in their study reported flashbacks following VR exposure.

In response to reports of prolonged and delayed aftereffects, the military instituted mandatory grounding policies for post-simulator flights in order to guard against the negative aftereffects that can occur subsequent to training in a flight simulator (Crowley, 1987; Kennedy, Lane, Lilienthal, Berbaum, & Hettinger, 1992). A simulator sickness field manual, developed by the U.S. Department of Defense and distributed to all military simulator sites, stated that flight personnel should be grounded (i.e., flights should not be scheduled) for at least 24 hours after simulator exposure or 12 hours after simulator sickness symptoms have subsided, whichever is longer (Naval Training Systems Center, 1989). Obviously, restrictions on the post-simulator activities of flight personnel can affect operational readiness, but the military also recognized the potential risk to pilots as well as to the expensive equipment under their control (Kennedy, Hettinger, & Lilienthal, 1990). Recently, the Department of the Navy (2004) issued an update to the NATOPS (Naval Air Training and Operating Procedures Standardization) General Flight and Operating Instructions which included policy and procedural guidelines on simulator sickness. In addition to warnings about the occurrence of prolonged and delayed aftereffects, the aviation safety instructions also mandated that: (1) flight personnel experiencing simulator sickness abstain from flight duties on the day of simulator exposure and (2) flight personnel that have previously experienced simulator sickness cannot be scheduled for flight duty for at least 24 hours following exposure to a simulator.

Clearly, prolonged aftereffects, delayed effects, and flashbacks can present a significant threat to the afflicted user's activities for a considerable period of time following exposure. Kennedy and Stanney (1996) indicated that these types of long-term aftereffects occur in less than 10% of all flight simulator exposures. An overall incidence rate for VR systems has not been reported, although long-term aftereffects data from one VR study showed that 35% of

participants reported symptoms more than four hours after exposure and 17% reported symptoms the following morning (Stanney, Kingdon, & Kennedy, 2002). Kennedy and Stanney (1996) also suggested that, compared to flight simulators, the advanced technology in VR displays will produce "an even more serious level of impairment" (p. 61). Nevertheless, long-term aftereffects create the potential for the legal liability of VE designers, manufacturers, and system owners if an accident occurs as a result of VE exposure. It has been suggested that disorientation-type aftereffects such as dizziness have the greatest potential for causing personal injury (Baltzley et al., 1989). For example, symptoms of disorientation could compromise user safety while exiting the simulator (e.g., falling off of the stairs/ramp that must be traversed in order to leave the device). Disorientation, drowsiness, fatigue, and nausea, which are frequently reported following exposure to VE systems, can also affect an individual's ability to safely perform routine tasks such as walking, riding a bicycle, or operating a motorized vehicle (Kennedy, Kennedy, & Bartlett, 2002). If an accident occurs after the user is released from the VE facility and the cause can be associated with the aftereffects of VE exposure, the manufacturer or company that owns the VE device could be found legally liable and thus, required to pay compensation for damages (Kennedy, Kennedy, & Bartlett, 2002). At a minimum, the manufacturer or company could face costly and time-consuming litigation in order to defend a product liability claim.

Classifications of Sickness from Exposure to Provocative Environments

Motion sickness is a general term for the adverse signs and symptoms that are provoked exclusively or primarily by exposure to certain types of real or apparent motion (Money, 1970; Reason, 1969). The most frequently reported signs, or overt indications, of motion sickness are

vomiting, cold sweating, and pallor (whitish-green skin hue), and the primary symptom, if the stimulus is sufficiently provocative, is nausea (Kennedy & Frank, 1985; Money, 1970). Other signs and symptoms that are considered reliable indicators of motion sickness are a general feeling of illness (malaise), headache, fatigue, and drowsiness (Harm, 1990; Kennedy & Frank, 1985; Money, 1970). Of course motion sickness is not a "sickness" in the usual sense of the term, but instead is the body's normal response to certain types of motion stimuli (Lawson, Graeber, Mead, & Muth, 2002; Money, Lackner, & Cheung, 1996; Reason & Brand, 1975). Money (1970) considered the term "motion sickness" inappropriate because it implies that vomiting in response to certain motions is unusual or abnormal. It is generally agreed that everyone (i.e., all people with a functioning vestibular system) can experience motion sickness provided that there is an appropriate stimulus of sufficient intensity and duration (Harm, 1990; Harm, 2002; Money, 1970; Reason, 1969). Accordingly, an absence of motion sickness in response to an extremely provocative stimulus would be indicative of a problem with the vestibular system (Lawson, Graeber, Mead, & Muth, 2002; Reason & Brand, 1975).

Motion sickness can be caused by exposure to a wide variety of motion environments. Generally, different types of motion sickness are named according to the particular environment in which the sickness was experienced. The oldest recorded form of motion sickness, which occurs in ships or boats, is aptly referred to as seasickness (Griffin, 1991; Money, 1970). Other common forms of motion sickness are associated with passive transport in different types of vehicles. These other variants of motion sickness are elicited by riding on certain carnival rides (e.g., swings), riding in an automobile (car sickness), bus, tank, train (train sickness), airplane or helicopter (airsickness), and during space flight which is termed space motion sickness (Förstberg & Ledin, 1996; Kennedy, Drexler, Stanney, & Harm, 1997; Reschke, 1990; Stott,

1990). It is interesting to note, however, that riding on a motorcycle does not produce motion sickness (Money, 1970). Similarly, the movements experienced while riding on camels or elephants can cause motion sickness, whereas motion sickness has never been reported from riding on horses (Money, 1970; Reason & Brand, 1975). Guignard and McCauley (1990) suggested that the stimulus for motion sickness while riding on these animals is the swaying or lurching gait of the camel and elephant, which is not found with horses. Visual stimulation without inertial motion (e.g., a wide-screen theater) is also sufficient to produce motion sickness, provided that the stimulus is the type that would normally be accompanied by vestibular and/or proprioceptive motion stimuli (Reason, 1969; Kennedy & Frank, 1985; Kennedy, Hettinger, & Lilienthal, 1990). Although, Money (1970) suggested that the signs and symptoms provoked by motion of the visual field are less severe (e.g., vomiting is rarely reported) than the sickness that occurs with movement of the body.

More recent manifestations of motion sickness occurred when individuals were exposed to virtual environment (VE) devices that used computer-generated imagery to create realistic, dynamic artificial environments (Kennedy, Drexler, Stanney, & Harm, 1997). These types of systems have the capability to simulate motion through changes to the visual imagery as the user moves within the synthetic environment. One type of VE device is a simulator that presents two-dimensional computer-generated images on a fixed-screen display (e.g., CRT, dome) and is typically used to simulate a flying or driving environment. The other type of device is a VR system, which employs a visually-coupled device (e.g., an HMD) that is worn by the user to present three-dimensional, computer-generated images. In addition to motion cues provided by changes in the visual scene, some simulators have a motion-base, synchronized with the video image, that provides concomitant physical motion (Guignard & McCauley, 1990).

Simulator sickness is a term used to describe the symptoms experienced by users during and/or after exposure to a simulator (Kennedy, Hettinger, & Lilienthal, 1990; McCauley, 1984). The first reports of simulator sickness occurred in the early 1960's, but technological deficiencies at the time inhibited further development of flight simulators (Kennedy, 1996; Kennedy, Drexler, & Compton, 1997). As computer technology advanced and became less expensive in the late 1970's, the U.S. military acknowledged the potential of flight simulators as a cost-effective training device by acquiring several flight simulators (Kennedy, Drexler, & Compton, 1997). Subsequent to fielding the newly acquired technology, reports of simulator sickness began to appear in nearly all of the military simulators including the Navy, Marine Corps, and Army (Crowley, 1987; Gower, Lilienthal, Kennedy, & Fowlkes, 1987; Kennedy, Lilienthal, Berbaum, Baltzley, & McCauley, 1989), Air Force (Warner, Serfoss, Baruch, & Hubbard, 1993), and Coast Guard (Ungs, 1987, 1988).

The results from extensive research on the side effects of exposure to military flight simulators have indicated that simulator sickness includes many of the signs and symptoms typically associated with motion sickness (e.g., nausea, sweating, pallor), and other symptoms such as disorientation, eyestrain, and dizziness which are not characteristic of 'true' motion sickness (Kennedy, Hettinger, & Lilienthal, 1990; Kennedy, Lilienthal, Berbaum, Baltzley, & McCauley, 1989). In addition to identifying differences in the patterns of symptoms, the research by Kennedy and his colleagues revealed that the symptoms which are similar to those of traditional motion sickness tend to be less severe and affect a smaller proportion of the exposed population (Kennedy, Hettinger, & Lilienthal, 1990; Kennedy, Lane, Berbaum, & Lilienthal, 1993). Based on these findings, the authors asserted that simulator sickness is distinctive from classical motion sickness.

McCauley and Sharkey (1992) referred to the artificial environments created by simulators and VR systems as 'cyberspace'. They argued that simulators are a particular type of VE system and as such, simulator sickness is actually a subset of motion sickness caused by VE exposure. Accordingly, McCauley and Sharkey proposed the use of a more general term, "cybersickness", for the symptoms provoked by exposure to both simulators and VR systems. Despite the authors' intention for their new term to represent symptoms from all types of VEs, a review of the literature suggests that many investigators have adopted the term "cybersickness" to refer specifically to VR systems. In general, it appears that most authors refer to the side effects of exposure to VR devices as cybersickness, or virtual reality sickness, in order to distinguish them from simulator-induced symptoms, where the term simulator sickness is still typically used. It should be noted that there are other investigators, though, who use the generic "motion sickness" term to refer to the side effects of exposure to any of the simulated environment systems (e.g., Durlach & Mavor, 1995).

Kennedy, Drexler, Stanney, and Harm (1997) pointed out that since the engineering goals of simulators were similar to those of other types of VEs, the problems with simulator sickness were expected to generalize to other VEs including VR devices that employ helmet-mounted displays (HMDs). While studies on flight simulators have shown that simulator sickness exhibits more oculomotor-related symptoms than conventional motion sickness, research on VR systems indicates that cybersickness exhibits more disorientation-related symptoms (Kennedy, Dunlap, Jones, & Stanney, 1996; Kennedy, Lane, Lilienthal, Berbaum, & Hettinger, 1992). Moreover, investigations into the motion sickness-like symptoms related to VR exposure have shown that these systems, especially those with an HMD, produce more severe levels of sickness than the sickness reported from exposure to flight simulators (Kennedy, Dunlap, Jones, & Stanney,

1996). In a survey of simulator sickness in ten different military flight simulators, approximately 10-60% of pilots reported some degree of sickness associated with exposure to the simulator (Kennedy, Hettinger, & Lilienthal, 1990; Kennedy, Lilienthal, Berbaum, Baltzley, & McCauley, 1989). In contrast, Kennedy, Jones, Stanney, Ritter, and Drexler (1996) found that the average level of sickness in their VR studies was not only significantly higher than those found in the flight simulators, but 85-95% of the study participants reported experiencing sickness symptoms. Similarly, Stanney, Kingdon, and Kennedy (2002) reported 88% of study participants reported symptoms immediately after exposure to a VR system.

Several investigators have proposed that the motion sickness provoked in one motion environment cannot be predicted from the sickness elicited in a different provocative environment. Kennedy, Dunlap, and Fowlkes (1990) cited a study by Thornton Linder, Moore, and Pool (1987) which found that the symptoms of space sickness were significantly different from the symptoms of classical motion sickness. Kennedy et al. (1990) therefore suggested that the types of symptoms produced in a particular environment may depend on the nature of the provocative stimulus. Similarly, Reschke (1990) stated that "each motion environment may have a similar but unique set of traits that distinguishes sickness in that environment" (p. 264). The scientific literature on simulator sickness and cybersickness appear to support this hypothesis. Simulator sickness and cybersickness typically involve visually-induced motion stimuli as opposed to traditional forms of motion-induced sickness that are caused by inertial motion. Also, vomiting is one of the cardinal symptoms of motion sickness and is very common in sea sickness, but is relatively rare in simulator sickness and cybersickness (Kennedy, Drexler, & Compton, 1997; Kennedy, Graybiel, McDonough, & Beckwith, 1968; Kennedy, Hettinger, &

Lilienthal, 1990; Kennedy, Lilienthal, Berbaum, Baltzley, & McCauley, 1989; Kingdon, Stanney, & Kennedy, 2001; Reason & Brand, 1975; Stanney, Kennedy, & Drexler, 1997).

Motion Sickness Theories

To date, there are no theories that have been developed to specifically address sickness in virtual environment systems. Instead, general theories of motion sickness have typically been applied to the study of these computer-generated systems in an attempt to identify the cause(s) of sickness provoked during or after exposure to a VE device. A number of considerably different theories on the nature of motion sickness have been proposed since the 1940's (Kennedy & Frank, 1985). Some of the older and generally unsupported theories of motion sickness (e.g., overstimulation, fluid shift, fear/anxiety) were reviewed in Kennedy and Frank (1985) and Reason and Brand (1975). The three major theories of motion sickness (sensory conflict theory, evolutionary theory, and postural instability theory) reviewed in the following sections focus on the interaction of the physical stimuli of the motion environment and the body's sensory systems, a physiological poison response mechanism, or the control of postural stability. The purpose of the review is not to critique the theories, but to provide the reader with an understanding of the discrete models which have been postulated to explain motion sickness in provocative environments. While each of the theories attempt to explain why motion sickness is provoked, Förstberg and Ledin (1996) pointed out that none of them have been particularly successful at formulating an overall motion sickness hypothesis. As a result, there is still no generally accepted theory that can satisfactorily account for, nor predict, all of the incidences of motion sickness, including sickness related to simulator and VR exposure.

Sensory Conflict Theory

Currently, the most widely accepted theory of motion sickness in real and virtual motion environments is the sensory conflict theory, which has also been referred to as the perceptual conflict, sensory rearrangement, sensory mismatch, neural mismatch, or cue conflict theory (Cheung, Howard, & Money, 1991; Harm, 1990; Harm, 2002; Hettinger, 2002). While James Reason (Reason, 1969, 1978; Reason & Brand, 1975) is typically credited with the first modern formulation of the sensory conflict theory, the premise was initially proposed by Irwin in 1881 in connection with seasickness (as cited in Förstberg & Ledin, 1996; Reason, 1978; Stott, 1990). Reason (1969, 1978) in fact stated that the sensory conflict model "makes no claim to originality" (p. 31 and p. 823, respectively). Reason (1978) indicated that the rationale behind the theory was to "define the essential nature of the provocative stimulus" (p. 819) of motion sickness by identifying the common sensory characteristics that provoked sickness in a variety of situations (e.g., seasickness, airsickness, etc.). Reason and Brand (1975) also believed that the most important contribution of the sensory conflict theory was to shift the focus of research at that time away from only vestibular aspects of sickness inducing situations (e.g., overstimulation theory) and toward identification of the type of information signaled by all of the body's position and motion receptors.

There are two premises of the sensory conflict theory. One assumption of the theory was that under normal conditions of movement, the visual, vestibular, and proprioceptive (somatosensory) systems simultaneously transmited correlated (i.e., redundant) information about the orientation and movement of the body (Reason & Brand, 1975). Accordingly, the central premise of the sensory conflict theory was that motion sickness occurred in situations where motion information cues received by these sensory systems were at variance with one

another *and* with the sensory input that the body (i.e., the central nervous system) expected to receive based on previous sensory-motor experiences (Cheung, Howard, & Money, 1991; Reason, 1978; Reason & Brand, 1975). In other words, motion sickness resulted from inconsistent information between the observed and expected motion cues (Förstberg & Ledin, 1996). Thus, Reason and Brand (1975) proposed that a provocative motion stimulus (i.e., one that produced motion sickness) always involved a conflict between the current spatial information and the information stored from prior experience.

The second premise of the sensory conflict theory was that for motion sickness to occur, the vestibular system must be one of the senses involved in the conflict, either directly or indirectly (Reason, 1978; Reason & Brand, 1975). Reason and Brand (1975) identified two major types of sensory conflict, or mismatches, that could produce sickness; an inter-modality conflict (e.g., between sensory receptors) and an intra-modality conflict (i.e., within a sensory system). The types of conflicts that could occur between the sensory systems included visual/vestibular and vestibular/proprioceptor (Förstberg & Ledin, 1996). Examples of the types of conflicts within a given sensory system included a vestibular-vestibular conflict, which was a conflict between the semicircular canals and otoliths contained within the vestibular apparatus (Kennedy & Frank, 1985), or a visual-visual conflict when there was a conflict between the focal and ambient visual systems (McCauley, 1984). Reason and Brand (1975), however, mentioned that canal-otolith conflicts were usually exacerbated by the presence of conflicting visual information. Kennedy and Frank (1985) also suggested that information from the visual system would have more salience than vestibular or proprioceptive cues due to sensory sensitivity and past history.

Reason and Brand (1975) presented the results from several different studies as support for the sensory conflict theory. One of the examples, which was presented as direct support for both premises of the theory, was a study by Parker (1971) in which participants watched a movie filmed from the inside of a car driving down a winding road. Observers who watched the film played in the normal (forward) mode became motion sick, which Reason and Brand suggested was due to the presence of visual motion in the movie without the corresponding vestibular cues that would be expected, based on past experience, in an actual vehicle. Interestingly, when observers watched the same film played backwards, motion sickness did not occur. Reason and Brand explained that, in this case, the combination of visual and (lack of) vestibular cues would not have contradicted the observers expectations since that particular set of cues never would have been encountered in an actual vehicle.

Virtual environment systems provide a highly visual world in which information is presented to the visual system, which can produce the perception of motion, but motion cues are typically not provided to the vestibular and proprioceptive senses (Biocca, 1992). According to the sensory conflict theory, the sickness from exposure to fixed-base simulators or virtual reality devices would occur because the visual stimuli provided by the device (i.e., apparent motion) were in disagreement with the vestibular and proprioceptive input that indicated the body was stationary. Similarly, in a moving-base simulator, sickness would result from the inability to resolve conflicts between the visual and inertial motion cues provided by the system and/or the stimuli did not match the users expectations based on their previous experience. Because the sensory conflict theory posited that the vestibular system, which only responds to angular and linear accelerations, had a vital role in motion sickness causation, Reason and Brand also suggested that an effective motion stimulus (real or illusory) must contain a changing velocity

component. For example, in a visually-induced sickness situation, the visual scene must imply a change in the direction or speed of the observer relative to the environment (Reason & Brand, 1975).

Kennedy and Frank (1985) and Harm (1990) pointed out that the sensory conflict theory has been used to explain the occurrence of motion sickness in a variety of motion environments (e.g., seasickness and space sickness). However, problems with the theory have been cited by several authors. One criticism was the theory's inability to account for a lack of sickness in situations where there was obvious sensory conflict (e.g., in static tilted rooms), and conversely, the occurrence of sickness in situations where there was little or no sensory conflict (Kennedy & Frank, 1985; Money, 1970). Kennedy and Frank (1985) criticized the model for not providing a way to predict the magnitude of the conflict (i.e., the severity of sickness) for a specific combination of sensory conflicts (e.g., visual-visual, visual-vestibular, visual-vestibular-proprioceptive). Money (1970) also highlighted the fact that bilateral labyrinthine defective subjects (i.e., those without a functioning vestibular apparatus) do not experience motion sickness, which contradicts the sensory conflict theory's hypothesis regarding the central involvement of the vestibular system. Cheung, Howard, and Money (1991) noted the theory's inability to explain why sensory input conflicts would manifest into symptoms of motion sickness. Other criticisms, summarized by Hettinger (2002), included an inability to determine "why the same conflict might not reliably produce sickness across different individuals" and "how to attempt to quantify the amount of conflict present in a given situation and relate it to the frequency and severity of motion sickness" (p. 483). Furthermore, Harm (1990) stated that the sensory conflict theory was not effective for explaining motion sickness from a physiological perspective nor was it able to predict motion sickness.

The most extensive critique of the sensory conflict theory was provided by Stoffregen and Riccio (1991). While the authors criticized the sensory conflict theory literature for the lack of explicit definitions for terms such as conflict, mismatch, and matching, their primary criticism was the 'expectation' component of the model. Specifically, Stoffregen and Riccio (1991) pointed out that a general principle of sensory conflict theory was that all conflict involved a violation of sensory expectations, but the authors believed that expectation violations should normally occur in any novel situation. The authors also noted that the conflict theory determined 'conflict' with reference to expectations of sensory cues, thus measurement of conflict depended on knowledge of an individual's experienced-based expectations. However, because there was no objective standard for an individual's expectations, Stoffregen and Riccio stated that an objective measure of conflict was not possible. Relatedly, they criticized the theory for the inability to predict, a priori, what information was being compared to an individual's expectations as well as the location of where the comparison would occur. Stoffregen and Riccio further argued that even if an expectancy violation existed, such a violation would not be sufficient to cause motion sickness.

Another major criticism by Stoffregen and Riccio (1991) was that the sensory conflict theory did not provide a basis on which to distinguish between situations that produced nausea and those which did not (i.e., provocative and nonprovocative situations). The authors explained that when sensory cues are redundant, the different sensory systems (e.g., visual, vestibular, and somatosensory) provide analogous information concerning the body's motion and/or spatial orientation and provide veridical information about the body's interaction with the environment. Stoffregen and Riccio stated that common implications in the sensory conflict literature were that: (1) the redundancy of information among different sensory systems served as the

"expectation" against which current sensory input was compared; and (2) an incongruence (i.e., nonredundancy) among the sensory signals produced sensory conflict. However, they challenged the theory's assumption that motion sickness was caused by nonredundant stimulation of the sensory systems and provided examples of situations where some level of conflict (i.e., nonredundancy) was present without producing symptoms of motion sickness.

In contrast to the sensory conflict theory, which assumed sensory cue redundancy was common, Stoffregen and Riccio (1991) claimed that redundancy of sensory system stimulation was very rare and nonredundant sensory information was actually common in natural and artificial environments including many nonprovocative situations. For example, they noted that during actual acceleration there is a normal nonredundancy between visual and gravitoinertial cues, but acceleration does not produce motion sickness. Accordingly, the authors asserted that past experience should not produce an expectation of redundancy across sensory systems and thus, redundant stimulation within and across sensory modalities could not serve as a criterion for conflict. Instead, they proposed that nonredundancy across stimulation of multiple sensory systems (visual, vestibular, and somatosensory) was relevant to perception and control of the body because it enabled adaptive changes in the control of behavior.

Finally, Stoffregen and Riccio (1991) contended that because the sensory conflict theory did not provide a basis on which to suggest that nonredundant sensory cues should be interpreted as sensory conflict in some situations and not in others, it did not provide a theoretical explanation for the existence of motion sickness. Stoffregen and Riccio (1991) further argued that without an independent basis for distinguishing conflict situations from other nonredundant situations, the conflict 'theory' becomes essentially a circular definition, "there is motion sickness because there is sensory conflict, and there is sensory conflict because there is motion

sickness" (p. 183). The authors proposed an alternative explanation of motion sickness, the postural instability theory, which is addressed in a subsequent section.

Evolutionary Theory

Treisman (1977) also criticized the sensory conflict theory and consequently proposed an explanation for motion sickness in terms of an evolutionary development. Money and Cheung (1983) noted that the occurrence of motion sickness, particularly nausea and vomiting, seemed to directly contradict evolutionary development because such an extreme adverse response to motion would not improve a species' survival. In fact, Förstberg and Ledin (1996) indicated that motion sickness for a person in a lifeboat at sea would greatly decrease their chance for survival. However, Treisman (1977) suggested there were mechanisms in the body that were responsible for initiating vomiting in order to purge ingested toxins from the body and thus, contributed to the survival of a species by eliminating the poison (Money & Cheung, 1983). Normally, ingested poisons affect the inner ear (i.e., the vestibular apparatus) causing a conflict between the cues from the vestibular and visual systems which signal the body that a poison was ingested and subsequently trigger a vomiting response (Kennedy & Frank 1985). Therefore, the vestibular mechanism that functions in the vomiting response to ingested poisons evolved as a biological protective mechanism (Money, Lackner, & Cheung, 1996).

Treisman's evolutionary theory (1977), sometimes called the poison theory, and the sensory conflict theory both support the idea that the body senses real or apparent motion through the visual, vestibular, and proprioceptive systems and that the signals received from these systems are continuously compared and calibrated with one another. But, Treisman's evolutionary theory differs in the supposition that the interaction of these sensory systems

evolved into a detection mechanism that indicated the presence of different types of toxins in the body (Förstberg & Ledin, 1996). The theory postulates that motion sickness, which ultimately leads to vomiting, is caused by a lack of correspondence among the signals received by the visual, vestibular, and proprioceptive sensory systems which subsequently stimulated the mechanisms of the vestibular system that normally facilitate the vomiting response to poisons (Förstberg & Ledin, 1996; McCauley, 1984; Money, 1970; Stoffregen & Riccio, 1991). In other words, motion sickness is the result of an erroneous interpretation that the motion-induced inconsistency between sensory cues are due to ingested toxins rather than the motion and "as a result the body inappropriately inflicts on itself" the signs and symptoms of motion sickness (Money, Lackner, & Cheung, 1996, p. 153). So, the vomiting that occurred with motion sickness was merely the body's attempt to eliminate toxins and the nausea would cause an aversion to the stimulus (Money, Lackner, & Cheung, 1996; Yardley, 1992).

Money and Cheung (1983) stated that the vestibular apparatus of the inner ear is primarily a sensory receptor for motion and gravity and results from empirical research have suggested that it plays an important role in the vomiting response to certain poisons as well as to certain motions. For example, it has been shown that bilateral loss of the vestibular apparatus prevents motion sickness, especially vomiting in response to motion (Money, 1970; Kennedy & Frank, 1985). Kennedy, Graybiel, McDonough, and Beckwith (1968) studied a group of bilateral labyrinthine defectives (LDs) aboard a ship in the North Atlantic and discovered that under storm conditions, all persons on the ship got sick except the LDs. Money, Lackner, and Cheung (1996) pointed out that LDs also are not susceptible to motion sickness related solely to visually induced motion. As mentioned previously, critics of the sensory conflict theory claimed that the theory could not explain why LDs do not experience motion sickness (Förstberg &

Ledin, 1996). In contrast, proponents of the evolutionary theory suggested that the finding of motion sickness immunity in LD individuals and animals supported the evolutionary theory because it provided evidence of the vestibular apparatus' involvement in the vomiting response (Crampton, 1990; Kennedy & Frank, 1985; McCauley, 1984; Money, 1970; Money, 1990; Money, Lackner, & Cheung, 1996; Reason & Brand, 1975).

Money and Cheung (1982, 1983) hypothesized that if the mechanisms responsible for motion sickness (i.e., the vestibular apparatus) function in response to poisons, then surgical removal of those mechanisms should affect the ability to respond to poisons. In order to determine whether loss of the vestibular apparatus could prevent, or at least impair, an emetic (vomiting) response to poisons, they tested the response of experimental animals to four different emetic poisons before and after bilateral surgical removal of the vestibular apparatus of the inner ear. Their results showed that after surgery the emetic response to certain poisons was impaired. Money and Cheung (1982, 1983) concluded that their experiment provided strong evidence to support the idea that the vestibular apparatus was part of the normal mechanism that facilitated the emetic response to certain poisons (cf. also Money, 1990). The authors also indicated that both Treisman's evolutionary theory and the sensory conflict theory of motion sickness are correct. They suggested that in provocative motion situations, the vestibular system reported conflicting sensory information to the brain which consequently required recalibrations between the visual and vestibular systems and a similar situation occurred when poisons were ingested. That is, excessive demands for recalibrations between these sensory systems were produced by conflicting information that was 'normally' received by the brain after the ingestion of poisons. Therefore, when conflicting information was created by exposure to motion, the brain interpreted

the unusual demands for recalibration as the result of ingested poisons, and thus initiated the protective emetic response (i.e., vomiting).

Relatedly, Money, Lackner, and Cheung (1996) hypothesized that if a poison response mechanism was the cause of motion sickness, as hypothesized by the evolutionary theory of motion sickness, then people who were more susceptible to emetic toxins should also be more susceptible to motion sickness. To support their hypothesis, the authors cited a study by Morrow (1985), in which chemotherapy drugs were administered to cancer patients. Morrow's study showed that individuals who reported themselves to be more susceptible to motion sickness also experienced more frequent, severe, and longer-lasting nausea and vomiting related to the chemotherapy drugs than the cancer patient control group that reported no history of motion sickness (Money, Lackner, & Cheung, 1996). Another finding cited by Förstberg and Ledin (1996) as support for the evolutionary theory of motion sickness was that infants (i.e., under two years old) are not susceptible to motion sickness; they are typically fed milk, which is not likely to be toxic, and they are usually exposed to sudden and unpredictable movements while being carried around. Money (1990) suggested that research showing many other species are susceptible to motion sickness (e.g., dogs, cats, monkeys, horses, some birds, etc.) also lends support to the evolutionary theory of motion sickness.

Postural Instability Theory

The sensory conflict theory was criticized for a number of reasons, which were discussed previously. However, Stoffregen and Riccio (1991) suggested that despite all of the problems with the sensory conflict theory, it remained the most widely accepted model of motion sickness, in part, because a 'credible' alternative was not available. Consequently, Riccio and Stoffregen

(1991) proposed the postural instability theory of motion sickness, based on an ecological approach to the perception and control of orientation and self-motion (i.e., action), that focused on behavior rather than stimulation of the sensory systems. Riccio and Stoffregen explained that the ecological approach to perception and action views "the interaction between the animal and the environment [as] the fundamental unit of analysis; neither can be examined separately" (p. 199). In their view, postural control was fundamental to all perception-control interactions with the environment and postural stability was determined by the interaction of the characteristics of the environment and the control skills of the individual (i.e., the ability to maintain or reestablish postural stability in a given situation).

Riccio and Stoffregen (1991) postulated a causal relationship between prolonged postural instability and the symptoms of motion sickness in provocative situations. In order to establish a link between motion sickness and postural stability, the authors cited a wide range of situations where motion sickness was related to factors that should influence postural stability. They argued that prolonged postural instability was present in motion sickness situations, but not in other (nonprovocative) situations. Riccio and Stoffregen hypothesized that motion sickness was caused by prolonged postural instability and that motion sickness would occur in situations where an individual had not learned effective strategies to maintain postural stability. Stoffregen and Riccio also claimed that in some situations, an individual may be unwilling or unable to terminate their interaction with a provocative environment (e.g., riding in a car, boat, carnival ride, etc.) and as a result, prolonged postural instability may be present until adaptive control is achieved. Thus, they claimed that postural instability not only preceded motion sickness symptoms, but it was a necessary and sufficient condition to produce symptoms. Moreover, they alleged that the duration of instability would directly affect the likelihood and intensity of motion

sickness symptoms. Although, the authors noted that their theory did not account for the nature of motion sickness symptoms, only their existence.

Stoffregen and Riccio (1991) pointed out that in the sensory conflict theory, an individual's behavior has no causal role in motion sickness; it is merely one source of conflict. The authors rejected this view and asserted that self-controlled movement (e.g., head movement, control of the torso) does have causal significance in motion sickness. The authors proposed that provocative situations could be characterized by novel demands on the control of action (i.e., postural stability) as well as novel patterns in the stimulation of multiple sensory systems. They argued that the pattern of stimulation across sensory systems provided information about properties of the environment that influence the control of behavior. Thus, nonredundant patterns of stimulation across the sense organs provided complementary information, rather than conflicting information as suggested by the conflict theory, which resulted in adaptive changes in behaviors such as standing and walking (Riccio & Stoffregen, 1991). Hence, the postural instability theory suggested that in provocative situations, changes in sensory stimulation were determined by changes in how the environment constrained the control of posture (i.e., postural stability). Riccio and Stoffregen (1991) further suggested that when an individual is passively stable (e.g., lying down), information about postural stability is not relevant to behavior so postural control is not required. Therefore, they hypothesized that reductions in the incidence or severity of motion sickness should correspond to reductions in postural control demands such as closing the eyes or lying down as well as passive stabilization using seat-belts or head restraints. To support their argument, the authors cited several studies where passive restraint of the head dramatically reduced susceptibility to motion sickness.

Stoffregen and Riccio (1991) indicated that empirical studies involving postural stability typically only measured it before and after exposure to a provocative stimulus and measurement during stimulus exposure was rare. Postural instability (i.e., ataxia) has been reported as an aftereffect of exposure to VE systems (Kennedy, Berbaum, & Lilienthal, 1997; Kennedy, Drexler, Compton, 1997; Kennedy, Fowlkes, & Lilienthal, 1993; Kennedy & Stanney, 1996). However, according to Stoffregen and Riccio's (1991) theory, postural instability was the cause of motion sickness not just a symptom (or side effect) of exposure. They postulated that changes in postural stability subsequent to provocative stimulation were due to postural control strategies acquired during stimulus exposure. Relatedly, the sensory conflict theory was criticized for its inability to explain why LDs do not experience motion sickness. Therefore, Riccio and Stoffregen (1991) addressed the immunity of LDs to motion sickness in relation to the postural instability theory. First, they pointed out that the vestibular system was important to movement control, which was consistent with reports of reduced motor-control capabilities of LDs. Then, the authors suggested as a potential explanation for the finding that motion sickness was not induced in LDs was because they may behave differently from normal (i.e., vestibularly-intact) persons in provocative situations; LDs were able to adopt more stable control strategies in situations where others became unstable. Thus, Riccio and Stoffregen suggested that the LDs immunity to motion sickness was the result of changes in their postural control rather than the loss of the vestibular system, although they noted that there is no information about the patterns of movement in studies of LDs that could be used to support their hypothesis.

Stoffregen and his colleagues conducted several empirical studies on visually induced motion sickness in order to test their theory that motion sickness is caused by instability in the control of body posture (Smart, Stoffregen, & Bardy, 2002; Stoffregen, Hettinger, Haas, Roe, &

Smart, 2002; Stoffregen & Smart, 1998). Specifically, the authors indicated that the purpose of the studies was to identify an empirical relationship between visually induced motion sickness and postural instability and to determine whether postural instability preceded the onset of motion sickness symptoms. In these investigations, motion sickness and postural stability were assessed while standing participants were exposed to a moving room that provided an optical simulation of body sway (Stoffregen & Smart, 1998; Smart, Stoffregen, & Bardy, 2002) and seated participants were exposed to a fixed-base flight simulator (Stoffregen, Hettinger, Haas, Roe, & Smart, 2002).

In each of the visually induced motion sickness studies, participants were divided into two groups, Sick or Well, based on self-reports of motion sickness symptoms and the experimenter's judgment of observable symptoms (e.g., pallor) during stimulus exposure (Smart, Stoffregen, & Bardy, 2002; Stoffregen, Hettinger, Haas, Roe, & Smart, 2002; Stoffregen & Smart, 1998). The studies revealed significant differences between Sick and Well groups on a number of different postural stability measures (e.g., variability, velocity, and range of head movement) for both standing and seated participants and in both types of provocative environments. The results of the experiments showed that the Sick group exhibited more postural instability and the stability differences existed prior to the onset of motion sickness symptoms. The authors' conclusion in each of the studies was that the findings supported the central prediction of the postural instability theory: motion sickness was preceded by increased postural instability. However, caution must also be used in generalizing the results beyond these specific studies because all of the experiments employed small sample sizes ($n = 8$ to $n = 14$). Moreover, the number of participants in the 'Sick' group represented less than half of the sample size in each of the studies.

Riccio and Stoffregen (1991) also addressed the sickness related to simulator exposure in terms of the postural instability theory. They explained that in a simulator, control of the operator's body was constrained by the properties of the simulator, not of the vehicle simulated. Thus, motion sickness occurred in fixed-base simulators (or VR systems) because prolonged postural instability was induced by inappropriate postural adjustments in response to visually specified motions (i.e., accelerations and rotations) in the simulated environment. Relatedly, the postural instability theory predicted that motion sickness would not occur in situations where passive stability was achieved through full restraint because the demands on postural control would be eliminated, although they pointed out that complete restraint was not practical in the real-world (Riccio & Stoffregen, 1991). Therefore, the authors predicted that the incidence of motion sickness would be a function of the degree of passive restraint, particularly restraint of the head and torso, where more restraint would produce less sickness. Jones (1998) empirically tested this hypothesis by exposing two groups of participants, unrestrained and restrained (head, neck, and torso), to a fixed-base driving simulator. The results of the investigation revealed a significant difference in postural stability (lateral head movement), where the unrestrained group moved more than the restrained group, but there was no effect of restraint on the severity of sickness. Participants who moved more during stimulus exposure did not experience greater sickness, which contradicted the result predicted by the postural instability theory.

Warwick-Evans and Beaumont (1995) believed that both sensory conflict and postural instability were present in situations that provoked motion sickness. Hettinger (2002) also argued that the sensory conflict theory could be used to explain many of the situations that ultimately lead to prolonged disruptions of postural control. In order to simultaneously evaluate competing predictions of the sensory conflict and postural instability theories, Warwick-Evans

and Beaumont (1995) conducted an investigation that attempted to decouple sensory conflict from postural instability. In their study, postural stability and motion sickness were measured while participants were seated in a chair and watched a 20 minute video, taken from the viewpoint of a person walking around a college campus. The experimenters varied the level of sensory conflict by exposing one group to the film at the normal speed and a second group to the film at a 40% faster speed. Both groups were partially restrained (i.e., head restraint) to reduce postural instability and to control the level of instability across the two levels of sensory conflict.

Warwick-Evans and Beaumont (1995) found that symptoms of motion sickness were produced in both conflict conditions, but significantly faster symptom onset was found in the lower sensory conflict (normal speed) condition. Additionally, between-group differences in postural stability (i.e., movement frequency and magnitude) were found for the two conflict conditions. The investigators concluded that their results were inconsistent with those predicted by the postural instability theory, although they noted that the equipment used to measure movement may have limited detection of smaller movements that might have revealed postural control differences. Furthermore, while the authors stated that the postural instability theory was more ecologically valid than the sensory conflict theory, they also indicated that the current form of the theory was not empirically supported. Warwick-Evans, Symons, Fitch, and Burrows (1998) later conducted two similar studies using two levels of sensory conflict, but in these studies they also manipulated the level of postural restraint (free standing and lying down). The results of both experiments showed no significant difference in motion sickness symptoms between the two restraint conditions. Therefore, their findings contradicted the postural stability theory's claim that a reduction in postural control demands would reduce motion sickness.

Measurement of Sickness

A valid and reliable measure is required to assess and quantify the affects of simulator or VR exposure on the individuals exposed to the device. Various objective and subjective measures of motion sickness have been used to document the effects of exposure to different provocative environments. The following sections review the most common measurement techniques used to quantify motion sickness with an emphasis on sickness related to simulator and VR exposure. It should be noted that the use and meaning of the term "sickness" is very inconsistent in the scientific literature, which can create a great deal of confusion for the reader. Specifically, a review of the motion sickness literature revealed that authors have used the term to indicate the presence (or absence) of a wide range of overt signs and/or symptoms of motion sickness. For instance, some authors stated that individuals were motion sick when they reported only nausea. Other authors only used the term when individuals vomited during the study while other articles used "motion sickness" to refer to individuals that reported a constellation of signs and symptoms. Moreover, an author's definition of motion sickness is often ambiguous, merely reporting individuals as "sick" or "not sick".

Objective Measures

A variety of objective techniques have been developed in an effort to measure and record the signs and/or symptoms of motion sickness. The Motion Sickness Incidence (MSI) was the most simplistic objective measure of motion sickness. For the MSI measure, the number of individuals who vomit from exposure to a particular provocative stimulus were counted and the number was then expressed as a percentage of the total number of persons exposed to the stimulus (Wertheim, 1999). While the measure was straightforward, there were several

problems associated with use of the MSI. First, the MSI only assessed vomiting, so other effects that were not strong enough to elicit vomiting, but still potentially debilitating (e.g., severe nausea), were totally discounted in the assessment of motion sickness severity (Wertheim, 1999). Second, use of a dichotomous criterion (i.e., no vomit/vomit) statistically constrained the reliability of the sickness measure (Lane & Kennedy, 1988). Third, the number of people who vomit in simulator and VR studies is relatively low (i.e., less than 0.3% in simulators [Kennedy, Drexler, & Compton, 1997] and less than 2% in VR devices [Kingdon, Stanney, & Kennedy, 2001]) compared to the number of individuals who experience other symptoms of sickness. Therefore, a large number of participants would be needed to establish a valid MSI score, which is usually not feasible in VE studies (Wertheim, 1999). Finally, use of the MSI measure clearly required the provocative stimulus to be continued to the point of vomiting, which has obvious negative implications for obtaining Institutional Review Board (IRB) approval of a study as well as for participant recruitment and experiment attrition rates.

Other efforts to objectively measure motion sickness incidence and severity have focused on the development of physiological indices. The primary governmental agency involved in the development of physiological measures of motion sickness is NASA, which has had a major program of research dedicated to developing objective measures of sickness including performance (behavioral) measures for over 40 years (Kennedy, 1996). While the behavioral measures have not been particularly successful, measures of sensorimotor functions such as posture, vestibulo-ocular reflex, and past pointing have shown better results (Kennedy, 1996). Kennedy noted, however, that the expense and the non-portability of the equipment limits the use of these particular tests.

The effects of various motion sickness stimulus conditions on different physiological response variables have been reported in the scientific literature. The physiological parameters used in other motion sickness research have included the effects on the cardiovascular system (heart rate, blood pressure, pallor), respiratory system (volume, rate), gastrointestinal system (tone, motility), and various stress hormone levels in the neuroendocrine system (Harm, 1990, 2002; Kennedy & Frank, 1985). However, the development of valid and reliable objective measures to index motion sickness severity have generally not been successful. Reason and Brand (1975) reported that the general findings on changes accompanying motion sickness revealed that cardiovascular and respiratory measures were inconsistent and unreliable. More recent investigations of cardiovascular indices have reported similar difficulties. For example, Johnson, Sunahara, and Landolt (1993) evaluated changes in blood flow as a potential physiological index of motion sickness. The researchers found a statistically significant correlation between blood flow changes and nausea severity, but the effect was small and therefore considered an unreliable measure for individual subjects (Wertheim, 1999). Similarly, Wertheim reported that decreases in oxygen consumption, which were initially thought to be associated with motion sickness, were actually the result of reduced muscular activity.

Objective measures of pallor, a cardinal sign of motion sickness, have also been investigated using techniques to index blood volume in the skin such as infrared reflectance plethysmograph (i.e., palorimetry), developed by Oman and his colleagues, and transcutaneous oxygen level used by Harm (Harm, 1990, 2002; Kennedy, Fowlkes, & Hettinger, 1989). While these measures have shown a relationship between changes in skin blood flow and intensity of stomach-related symptoms, individual differences in the pattern of skin color changes were also observed (Harm, 1990, 2002). Another cardinal sign of motion sickness is cold sweating, or

sweating in the absence of a thermal stimulus (Reason & Brand, 1975). Warwick-Evans et al. (1987) empirically evaluated whether electrodermal activity (i.e., skin conductance related to sweating) could index the intensity of motion sickness. Their findings showed a consistent and positive association between increases in skin conductance and self-reports of motion sickness. However, the authors noted that the measure was overly sensitive to psychological (e.g., anxiety) and physiological (e.g., ambient temperature, motor-activity) influences. Reason and Brand (1975) also reported the sensitivity of pallor and cold sweating to factors other than motion stimuli (e.g., anxiety, stress) and as a result, declared that these signs by themselves could not be used to establish the existence of motion sickness.

The experimental evidence on gastrointestinal changes has suggested a relatively consistent reduction in gastric tone and motility accompanying motion sickness onset (Reason & Brand, 1975). Specifically, an increase in gastric motility called tachygastria has been empirically related to the onset of motion sickness symptoms (Kennedy, Hettinger, & Lilienthal, 1990). Using cutaneously-recorded (i.e., surface electrodes placed on the abdomen) electrogastrograms (EGGs), Stern found that tachygastria immediately preceded subjective reports of motion sickness (Stern, Hu, Anderson, Leibowitz, & Koch, 1990; Stern, Hu, Vasey, & Koch, 1989; Stern et al., 1985). Miller, Sharkey, Graham, and McCauley (1993) also found that physiological measures of skin conductance and tachygastria were sensitive to self-reports of simulator sickness. However, the authors noted that their analyses suggested physiological variables may predict motion sickness discomfort when it is restricted to within-subject comparisons, but not when combined across subjects. Similarly, Harm (1990) reported that when changes in physiological measures were averaged across susceptibility groups (not sick,

mildly sick, and severely sick), the results showed only small differences between groups, which suggested the physiological measures lacked the reliability necessary to predict sickness.

Although a wide range of physiological reactions in motion sickness have been observed in nearly every system of the body, there is still limited knowledge of the specific underlying physiological mechanisms responsible for the symptoms of motion sickness (Harm, 1990, 2002). The primary reason cited by the author was the lack of consistency reported in the literature on physiological responses to motion stimuli for almost all of the physiological variables examined. Harm (1990, 2002) noted an equal number of reports could be found where the physiological variable(s) under investigation increased, decreased, or did not change in response to a motion stimulus. Furthermore, the reported inconsistent findings applied to individual responses within a given study, within individuals exposed repeatedly to the same stimulus, as well as across different experiments (Harm, 1990). Harm (1990, 2002) suggested the inconsistent findings could have been due to individual subject differences, stimulus conditions, severity of sickness, the specific physiological measure used, or the methodology used to measure and analyze the particular physiological response. For instance, a wide variability in the number and complexity of individual reactions to provocative stimuli has been reported across different individuals and stimulus conditions in the type, severity, and time-course of physiological responses (Harm, 2002). Kennedy, Dunlap, and Fowlkes (1990) also cited uncontrolled factors in real-world motion sickness investigations, particular symptoms reported by individuals may depend on the nature of the provocative stimulus, and individual differences in symptom response patterns as potential contributing factors in the difficulty associated with the development of an accurate objective measure of motion sickness. For example, the number of symptoms experienced from exposure to a simulated environment can vary; some people exhibit all or several of the

symptoms while others exposed to the same device may only experience a few symptoms or no symptoms at all (Kennedy & Fowlkes, 1992). Moreover, large differences in susceptibility to motion sickness have been reported; some individuals may be totally incapacitated by exposure to a particular motion stimulus while others remain unaffected (Kennedy, Dunlap, & Fowlkes, 1990).

The ability to develop an objective measure of motion sickness has also been limited by the low reliability and/or insensitivity of the measures used in the investigations (Kennedy, 1996). Kennedy and Fowlkes (1992) suggested that all types of motion sickness, including sickness from simulator and VR exposure, involve multiple symptoms (i.e., motion sickness is polysymptomatic). Kennedy (1996) noted that the diversity of potential symptoms suggests there are numerous potential measures of sickness. Nevertheless, Kennedy and Fowlkes (1992) indicated that an assessment of only one sign or symptom could not provide a sensitive metric of motion sickness and as a result, would not offer any meaningful conclusions for the investigator (Kennedy & Fowlkes, 1992). Kennedy, Dunlap, and Fowlkes (1990) also stated that because motion sickness is a very complex phenomenon, "any single criterion will have substantial psychometric limitations" (p. 205). Accordingly, a measure of sickness induced by real or simulated motion *must* reflect the polysymptomatic nature of the syndrome (Kennedy & Fowlkes, 1992).

In contrast to objective measures which only evaluate a single sign or symptom of motion sickness, subjective measures such as self-report questionnaires typically assess multiple symptoms of motion sickness through the use of symptom lists. Motion sickness also often includes a variety of subjective symptoms such as fatigue, eyestrain, and drowsiness that cannot be objectively measured. Moreover, Yardley (1992) indicated that because individual

physiological measures have only shown a moderate covariance with subjective symptoms of motion sickness, self-report or observer-reported ratings of multiple symptoms are typically employed in motion sickness research. Consequently, most researchers use subjective methods, particularly self-report questionnaires, to measure and quantify the incidence and severity of motion sickness (Wertheim, 1999).

Subjective Measures

Reason and Brand (1975) supported the use of subjective data to index motion sickness because they considered subjective reactions as "the single most valuable source of information about the subject's condition" (p. 82). Lawson, Graeber, Mead, and Muth (2002) also acknowledged the importance of subjective reports of motion sickness and stated "a great deal can be learned by careful inquiry into the subjective aspects of motion discomfort" (p. 599). However, Wertheim (1999) mentioned an often cited concern of some investigators related to the use of subjective measures, that is, the validity of subjective data. Specifically, can self-report measures serve as a valid tool to quantify the incidence and severity of motion sickness? In response, the author cited several studies where the validity of self-report rating scales was established by showing that averaged group self-report ratings were highly correlated with averaged group MSI scores (i.e., an objective measure of motion sickness). Kennedy, Dunlap, and Fowlkes (1990) also pointed out that the validity of a dependent measure is limited by its reliability. But, in motion sickness research "no one symptom, regardless of how well it is measured, can be statistically reliable enough" (Kennedy, 1996, p. 30).

Lawson, Graeber, Mead, and Muth (2002) reported that subjective reports of symptoms of motion discomfort have been proven to be valid and reliable measures of an individual's

physical state and are also important criteria in the interpretation of physiological and performance effects of exposure to provocative stimuli. Thus, while many researchers prefer to use objective measures to evaluate a criterion under investigation, the choice of a motion sickness measure must be based on the reliability and validity of the measure rather than whether it provides objective or subjective data. Moreover, even the choice of an objective measure is ultimately a subjective decision (R.S. Kennedy, personal communication, November 4, 2004, based on comments by N.E. Lane after F. Muckler prior to 1981). Relatedly, based on a literature review of the methods available, Wertheim (1999) asserted that an investigator's preference for and choice of a measurement tool for assessing motion sickness symptoms often seemed arbitrary.

There are several advantages of self-report data including the ease of use, ability to collect a significant amount of information from participants in a short period of time, noninvasive measurement, and minimal cost (Kennedy, 1996; Kennedy, Jones, Stanney, Ritter, & Drexler, 1996). Additionally, self-report questionnaires typically include lists of symptoms which provide a more sensitive metric than the objective measurement of a single sign or symptoms (Kennedy & Fowlkes, 1992). However, it should be noted that a disadvantage of self-report data is the reliance on respondents willingness to truthfully respond to inquiries.

Self-Report Measures of Motion Sickness

Self-report questionnaires have been the primary technique used to measure and quantify the incidence and severity of motion sickness (Kennedy, Hettinger, & Lilienthal, 1990; Wertheim, 1999). For example, Kennedy (1996) estimated that 90% of the information on simulator sickness was derived from self-report questionnaires. In fact, most of the scientific

information on motion sickness, including sickness from exposure to simulated environments, was obtained with self-report questionnaires, which employed some type of weighting (or aggregate) procedure to form composite scores that were used to characterize the severity of sickness (Kennedy, 1996; Kennedy, Jones, Stanney, Ritter, & Drexler, 1996).

A brief review on the history of the development of self-report motion sickness questionnaires was provided by Kennedy, Drexler et al. (2003). The authors indicated that the earliest technique to use scaled values for scoring motion sickness was developed by Wendt in the mid-1940s. His technique employed a three-point rating scale to index the degree of motion sickness severity where 'vomiting' was assigned the highest score, followed by 'nausea without vomiting' and 'no symptoms' (Kennedy, Drexler et al., 2003). Birren subsequently adopted Wendt's rating scale technique for use in studies on seasickness (Lane & Kennedy, 1988). Then in 1960, Graybiel, Clark, and Zarriello developed the first multi-symptom checklist, which only contained seven symptoms, for studying motion sickness in the Pensacola Slow Rotation Room (Kennedy, Drexler et al., 2003). However, Kennedy, Drexler et al. (2003) reported that the self-report technique was not formalized until 1965 when Kennedy and Graybiel expanded the 7-item symptom checklist and created a new scoring procedure. Kennedy and Graybiel (1965) employed a protocol analysis technique to record participant's verbal reports of symptoms during Coriolis-induced sickness in the Slow Rotation Room. The authors then combined the verbal symptom reports with Graybiel's 7-item checklist, which resulted in a new 33-item symptom checklist, that was later named the Motion Sickness Questionnaire (MSQ). In order to quantify the symptoms, Kennedy and Graybiel created a five-point composite score, based on Wendt's original rating scheme, that provided an index of the overall severity of motion sickness (Kennedy, Drexler et al., 2003).

The Graybiel symptom rating scale, also known as the Graybiel classification system, was a subjective motion sickness measure that was based on the MSQ checklist. Like the MSQ, this rating scale contained a list of symptoms, however, the severity score was based on a combination of weighted experimenter and subject ratings of symptom severity (Lawson, Graeber, Mead, & Muth, 2002; Wertheim, 1999). Although the Graybiel measure was validated to some extent, Wertheim (1999) reported that several researchers questioned the assumptions of the metrics underlying the complex scoring method and as result, they adopted Graybiel's symptom list, but developed their own simplified method of scoring (see also Lawson et al., 2002). Other investigators selected symptoms found in the motion sickness literature and created their own symptom list and scoring methods (Flanagan, May, & Dobie, 2005; Gianaros, Muth, Mordkoff, Levine, & Stern, 2001; Miller & Muth, 2004; Wertheim, 1999). Wertheim pointed out, however, that these individually created measures paid little attention to validity issues; the basis for symptom selection was often unclear or not reported by the author and none of the symptom lists were validated. A notable exception was the Motion Sickness Questionnaire that was mentioned previously (Wertheim, 1999).

As stated previously, the Motion Sickness Questionnaire (MSQ), sometimes referred to as the Pensacola Motion Sickness Questionnaire, was originally developed almost 40 years ago by Navy scientists to assess and quantify subjective reports of motion sickness symptoms in their studies on various types of motion sickness (Kennedy & Graybiel, 1965; Kennedy, Jones, Stanney, Ritter, & Drexler, 1996; Lane & Kennedy, 1988). Development of the MSQ was based on extensive research employing data collected from a number of different motion environments (Lane & Kennedy, 1988). Early MSQ investigations used highly provocative stimuli which Lane and Kennedy (1988) reported were severe enough to induce vomiting, near-vomiting, or a

request for early termination of exposure in practically all of the study participants. The MSQ was used to collect symptom data in studies of: Coriolis sickness in a rotating room (Kennedy & Graybiel, 1965; Kennedy, Tolhurst, & Graybiel, 1965), weightless conditions (Kellogg, Kennedy, & Graybiel, 1965), seasickness aboard naval ships, airsickness in aircraft flying through hurricanes, space sickness in a series of NASA studies, as well as simulator sickness in several high-fidelity flight simulators (Lane & Kennedy, 1988; Lawson, Graeber, Mead, & Muth, 2002).

The MSQ consisted of a paper-and-pencil checklist of 33 separate major and minor symptoms typically associated with the onset of motion sickness (e.g., nausea, headache, apathy; Lane & Kennedy, 1988). However, depending on the study in which it was used, the number of symptoms included in the checklist could vary from 20 to 33 (Kennedy, Drexler et al., 2003). Nevertheless, Lane and Kennedy (1988) declared that the largest number of symptoms which were appropriate to the type of sickness under investigation should be used to enhance the reliability of the measure because larger numbers would provide symptom redundancy. A larger number of relevant symptoms would also ensure that all of the important dimensions of the particular type of sickness under investigation would be represented within the symptom checklist (Lane & Kennedy, 1988).

Each symptom on the MSQ checklist was rated by the participant in terms of the degree of severity on a four-point ordinal scale with anchor points at 'None, Slight, Moderate, and Severe', although some of the symptoms required a 'yes/no' response (Kennedy, Jones, Stanney, Ritter, & Drexler, 1996). A diagnostic scoring procedure was then applied to the checklist which generated a composite, global sickness score that reflected the overall discomfort of the respondent (Kennedy, Jones, Lilienthal, & Harm, 1994). The global severity score ranged from

zero, indicating no symptoms were reported, to the highest score possible, indicating an individual vomited, but the highest score varied in different types of motion sickness studies (Lane & Kennedy, 1988). Lane and Kennedy asserted that the reason for converting subjective symptom reports into scaled numbers was to allow statistical analyses of the data for use in scientific research.

After its use in a number of simulator sickness investigations, Lane and Kennedy (1988) noted several deficiencies in the MSQ as a measurement device for simulator research. The major problem cited by the authors was that the scoring method provided a single global severity score, which would only be appropriate for studies concerned with the overall severity of sickness. However, Lane and Kennedy (1988) remarked that motion sickness was known to be multidimensional (i.e., produced a variety of symptoms), so a single numerical indicator might not provide the best diagnostic information that would be available from individual measures of the separable dimensions underlying motion sickness. The authors also noted the need for a reliable measure that could be used to assess symptoms produced in situations less severe than the conditions in the motion sickness studies which were used to develop the MSQ (i.e., testing to the point of vomiting).

Differences between traditional motion sickness and simulator sickness also suggested that the MSQ was not an ideal measure of simulator sickness (Lane & Kennedy, 1988). Specifically, simulator exposures produced symptoms similar to 'classic' motion sickness, but the symptoms were usually less severe than motion sickness and typically affected a smaller proportion of the exposed population (Kennedy, Lane, Berbaum, & Lilienthal, 1993). Also, some of the symptoms that were valid in the MSQ scoring method were not appropriate for measuring simulator sickness because they were rarely reported in simulator exposures (e.g.,

vomiting; Kennedy, Lane et al., 1993; Lane & Kennedy, 1988). Furthermore, Lane and Kennedy (1988) cited strong visual and visual-motion stimuli in simulator studies which were generally not present in other motion sickness situations. Consequently, the authors declared that simulator sickness was sufficiently different from motion sickness to justify the use of a separate measurement instrument specifically designed to quantify sickness related to simulator exposure. Lane and Kennedy (1988), therefore, reanalyzed the MSQ using factor analyses of flight simulator data which resulted in a modified version of the MSQ called the Simulator Sickness Questionnaire (SSQ) described below.

Simulator Sickness Questionnaire

A calibration sample of more than 1100 pairs of 28-item MSQ checklists (i.e., pre- and post-exposure), collected from ten different flight simulators, were reanalyzed (Lane & Kennedy, 1988). The authors' objective in reanalyzing the MSQ data was to develop the Simulator Sickness Questionnaire (SSQ) that would provide a more appropriate and valid index of overall severity for *simulator* sickness, diagnostic subscale scores that could offer information about the potentially separable dimensions of simulator sickness, and a more powerful and convenient scoring method (Kennedy, Lane, Berbaum, & Lilienthal, 1993).

Preliminary analyses focused on determining which MSQ symptoms were relevant for an index of simulator sickness (Lane & Kennedy, 1988). First the authors carefully reviewed the MSQ data in order to identify the symptoms that showed systematic changes from pre- to post-simulator exposure. Any symptoms reported less than 1% of the time, showing no change, or showing a decrease in severity or frequency were eliminated from further analyses. Lane and Kennedy (1988; Kennedy, Lane et al., 1993) expected some variability in symptom severity

among the devices because the MSQ data were collected from ten different simulators which were known to vary in the overall level of sickness severity. Accordingly, the MSQ data were also reviewed to identify any symptoms that exhibited different levels of severity or frequency across simulators, and those symptoms were selected for inclusion in the modified checklist (Kennedy, Lane et al., 1993; Lane & Kennedy, 1988). Based on the authors' analyses, 16 of the MSQ symptoms were ultimately retained as important indicators of simulator sickness.

The 16-item symptom SSQ checklist list was then factor analyzed in an attempt to extract reliable sickness subscale measures that could be used to provide information about the particular systems of the body which were affected by a provocative motion stimulus (cf. Lane and Kennedy, 1988 for a detailed description of the factor analysis procedures). The results of the factor analytic procedures revealed that the symptoms fell into three-, four-, five-, or six-factor solutions (i.e., symptom clusters). However, the three-factor solution was considered to be the most appropriate because the additional factor solutions did not contain a sufficient number of symptoms to provide reliable subscale scores (Kennedy, Lane et al., 1993; Lane & Kennedy, 1988). Moreover, Kennedy, Lane et al. (1993) reported that results from other factor analyses of MSQ data collected in related stimulus domains (e.g., prolonged visual display unit [VDU] use, seasickness) yielded similar symptom clusters.

The three factors, which formed the basis of the three SSQ subscales, were labeled Nausea, Visuomotor, and Disorientation (Lane & Kennedy, 1988). It is important to note that the Visuomotor factor was renamed Oculomotor in Kennedy, Lane et al. (1993) and is referred to as such in all subsequent publications related to the SSQ subscales. Scores on the Nausea (N) subscale, which were associated with the autonomic nervous system, represented symptoms related to gastrointestinal distress (e.g., nausea, stomach awareness, and burping; Kennedy, Lane

et al., 1993; Lane & Kennedy, 1988). Scores on the Oculomotor (O) subscale, reflected symptoms related to disturbances of the visual system and included symptoms associated with seeing (e.g., difficulty focusing) and visual fatigue (e.g., eyestrain, headache). Scores on the Disorientation (D) subscale were related to disturbances of the vestibular system (e.g., dizziness, vertigo). The authors argued that the three SSQ subscales represented different 'target' systems in the body that were affected by stimulus exposure. Thus, depending on the mechanisms affected, exposure to a given simulator could cause symptoms that appear in none, one or more, or all of the symptom clusters (Kennedy, Lane, Berbaum, & Lilienthal, 1993; Lane & Kennedy, 1988). The authors, therefore, maintained that the SSQ subscales could be used to identify "where and in what ways a simulator may be causing problems for the user" (Lane & Kennedy, 1988, p. 15; Kennedy, Lane et al., 1993, p. 208). In addition to the three subscales, the factor analysis revealed a global measure of overall sickness severity similar to the MSQ, known as the Total Severity (TS) score, that could be used as a general index of whether a particular device was producing a sickness problem (Lane & Kennedy, 1988).

Scoring Method

An important underlying assumption of the SSQ scoring method was that individuals which reported themselves as not in their usual state of fitness were excluded from analysis (Kennedy, Lane et al., 1993; Lane & Kennedy, 1988). Reports in the scientific literature have shown that illness (e.g., flu, cold, etc.) can increase an individual's susceptibility to motion sickness (DeWit, 1957; Kennedy, Hettinger, & Lilienthal, 1990; Wright, 1995). Therefore, a list of questions designed to assess an individual's current state of health was included in the pre-

exposure SSQ, which investigators could use to exclude 'unhealthy' participants from the sample (Kennedy, Lane et al., 1993; Lane & Kennedy, 1988).

The SSQ scoring method, developed by Lane and Kennedy (1988 Kennedy, Lane et al., 1993) and shown in Appendix A, used a weighting system to calculate scores with the following standard properties: (1) the lowest possible score on each subscale and the TS score was zero (i.e., no reported symptoms) and (2) a standard deviation of 15 for the scaled scores. The symptoms on the SSQ checklist, like the MSQ, were rated on a four-point ordinal scale anchored at 'None, Slight, Moderate, and Severe'. Accordingly, each symptom on the checklist was first assigned a value ranging from zero to three based on the severity of the rating: None = 0, Slight = 1, Moderate = 2, and Severe = 3 (Kennedy, Lane et al., 1993; Lane & Kennedy, 1988). Then, a score was computed for each subscale by summing the values of the symptoms corresponding to the particular subscale and multiplying that value by a specific unit weight (N = 9.54, O = 7.58, D = 13.92). Similarly, the Total Severity score was determined by summing the three unweighted subscale scores and multiplying by its unit weight (TS = 3.74). The authors stated that the function of the unit weights was to provide similar variabilities in the different scales which would enable a comparison of scores across the scales (Kennedy, Lane et al., 1993; Lane & Kennedy, 1988).

Validity and Reliability

Kennedy, Drexler et al. (2003) summarized the psychometric properties of the SSQ obtained from various motion sickness studies. The predictive validity of the SSQ was first reported in a seasickness study where the correlation between the SSQ Total Severity score and an objective measure of sickness (i.e., vomiting) was $r = 0.73$ ($p < 0.001$). Kingdon, Stanney,

and Kennedy (2001) also reported a significant correlation between participants who vomited during a VR study and scores on the SSQ Nausea subscale ($r = 0.65$, $p < 0.01$) as well as the SSQ Total Severity score ($r = 0.59$, $p < 0.05$). Results from simulator and VR studies have demonstrated that the SSQ is also a highly reliable measure. In a relatively large VR study (n = 200), Kennedy, Drexler et al. (2003) reported the SSQ split-half correlation was $r = 0.80$ and the correlation for the full SSQ, using Spearman's correction for test length, was $r = 0.89$. Similarly, the authors reported a reliability of $r \sim 0.78$ in a driving simulator study. Moreover, Kennedy, Drexler et al. (2003) indicated that research studies of motion sickness which employ an objective measure of sickness (e.g., physiological indices) often validate the measure against the score on a self-report questionnaire. Consequently, the authors argued that self-report measures such as the SSQ are "probably twice as reliable as the objective measures" that have been developed to replace them (p. 253).

Wertheim (1999) noted that the SSQ was the only validated instrument which could be used to measure the severity of simulator sickness. Based on his literature review of the methods available for assessing the magnitude of aftereffects, Wertheim specifically recommended use of the well-validated SSQ as an assessment tool "to obtain a more detailed and differentiated picture of the nature and severity of motion sickness simulator aftereffects" (p. 34). Similarly, Lawson et al. (2002) recommended use of the SSQ as a measurement tool for studies of sickness in simulated environments because unlike other self-report measures of motion sickness, the SSQ was specifically designed for use in less provocative environments (i.e., those with lower vomiting rates) as well as in situations that include some type of visual display. The authors also acknowledged the usefulness of the SSQ as a measurement instrument because it allows an assessment of the underlying symptom clusters (i.e., the SSQ subscale scores).

Factors Influencing Sickness in Virtual Environments

Early military flight simulators, which first called attention to the problem of simulator sickness, had equipment limitations such as visual distortions, excessive transport delays, and flickering images which were considered to be the source of the discomfort experienced by users (Drexler, Kennedy, & Compton, 2004). Simulator sickness was, therefore, initially thought to be due solely to the inadequacies of the equipment, so equipment improvements would eliminate the sickness problem (Kennedy, Jones, & Dunlap, 1996). However, as technological advances improved the fidelity of the equipment and the visual scenes became more realistic, the incidence and severity of sickness actually increased (Kennedy, Berbaum et al., 1987; Kennedy, Drexler et al., 2003; Kennedy & Lilienthal, 1994; Kennedy, Hettinger, & Lilienthal, 1990).

Kennedy and Fowlkes (1992) suggested that simulator sickness was not driven by a unitary cause, rather the source of the problem was a combination of factors (i.e., polygenic). Kennedy, Berbaum, Dunlap, and Smith (1995) argued that in order to control VE sickness, it was necessary to first determine which variables affected sickness and to what extent. Although the fundamental causes of motion sickness have not been completely identified, researchers have identified a number of factors that are thought to influence the incidence and severity of sickness related to VE exposure.

The first major effort to identify the causal factors of simulator sickness occurred in the early 1980's. In recognition of the importance of the problem, a three-day workshop on simulator sickness was convened by the National Research Council's Committee on Human Factors (McCauley, 1984). One of the main purposes of the workshop was to identify the likely cause of simulator sickness and the contributing factors. Participants of the workshop were all experts in their respective field (motion sickness, simulator sickness, vestibular dynamics, visual

processes, and simulator use and design). As a result of the meeting, a list of potential contributing factors of simulator sickness was generated. The factors included in the list focused primarily on simulator design characteristics, but a few operator characteristics were also identified (McCauley, 1984). These factors were grouped into the following five different categories: Motion and Vibration (frequency, acceleration, lags); Vision (field of view, display type, off-axis display); Visual Motion (refresh rate, temporal and spatial distortion, collimation); Simulator Features (motion/fixed-base, visual and motion system lags, washout); and Simulator Use (freeze, reset, seat position).

Since that initial groundbreaking meeting, other potential contributing factors to the sickness associated with exposure to simulator and VR systems have been identified. Furthermore, several investigators proposed a taxonomy for the different causal factors, although classification of the factors and the labels used for the determinant categories appeared to be highly subjective. For example, Kennedy and Fowlkes (1992) provided examples of several potential causal factors of simulator sickness and categorized them into three main types of determinants: Simulator Equipment Features, Simulator Usage, and Pilot Variables (i.e., state of fitness). Similarly, Kolasinski (1995) listed 40 factors as potential contributors to simulator sickness and grouped them into three major categories: Individual, Simulator, and Task factors. Although Kolasinski used different labels for the determinant categories (e.g., task factors instead of simulator usage), the constructs were the same as those in the Kennedy and Fowlkes taxonomy. In contrast, Kennedy and Fowlkes classified adaptation to sickness as a factor related to simulator usage whereas Kolasinski considered adaptation as experience with the system and therefore, classified it as an individual factor.

Another difference in the various proposed taxonomies of sickness determiners which suggested the subjective nature of the classification schemes involved the number and type of categories that were used to group the causal factors. For instance, as noted above, McCauley (1984) arranged the factors into five categories based on different aspects of the simulator whereas Kennedy and Fowlkes (1992) classified the factors into three categories based on the characteristics of the individual user, the equipment, and use of the equipment. Conversely, Kennedy, Berbaum, Dunlap, and Smith (1995) proposed a taxonomy of sickness determinants associated with exposure to VE systems which was composed of five major categories: Individual Differences, Equipment Features, Usage factors, Kinematics, and Duration. While the number of categories in their taxonomy was equivalent to those used by McCauley (1984), the nature of the categories was clearly different. Specifically, the Kennedy et al. (1995) taxonomy appeared to be an extension of the original Kennedy and Fowlkes (1992) taxonomy. A comparison of the two taxonomies revealed that the Equipment Features and Usage categories were retained in the Kennedy et al. taxonomy, but the Pilot Variable category was renamed to the more inclusive label, Individual Differences and the Kinematics and Duration categories were added.

Based on their taxonomy of sickness determinants and other findings from the literature on the main drivers of sickness in simulator and VR devices, Kennedy and his colleagues (Kennedy, 1996; Kennedy, Berbaum, Dunlap, & Smith, 1995; Kennedy & Smith, 1996) developed a preliminary causal model of sickness associated with exposure to simulated environments. The model, shown in Figure 1 below, contained the five major sickness determinant categories along with an estimation of the amount of variance accounted for by each element of the model (i.e., the relative importance of each category to the sickness criteria).

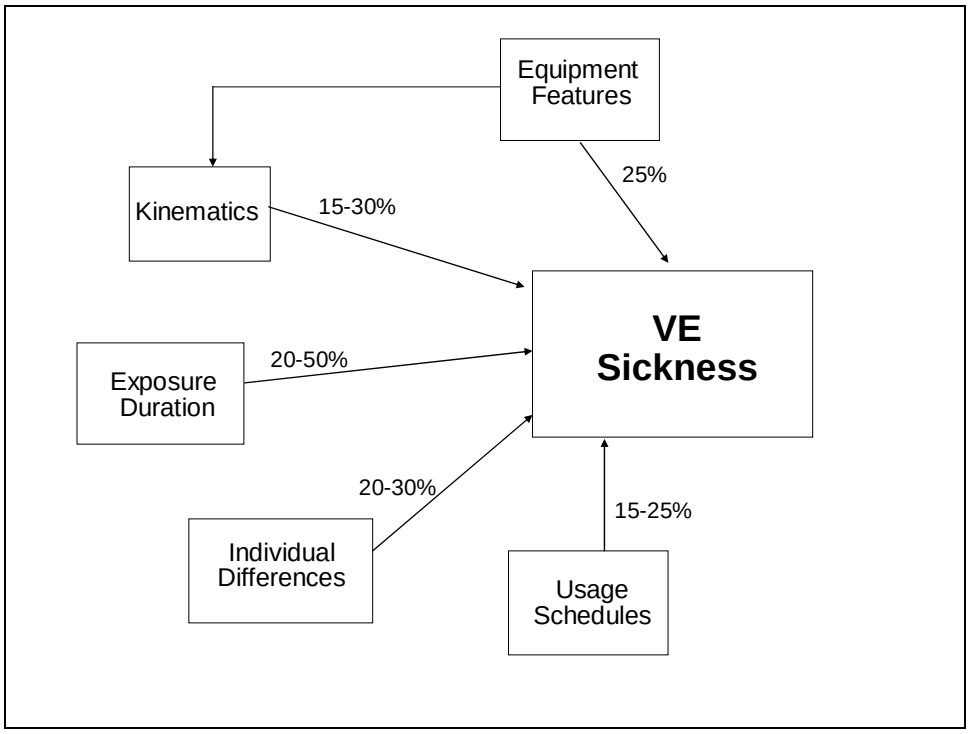

Figure 1. Model of the Potential Determiners of Sickness in Virtual Environments

Kennedy (1996) explained that the variance estimates for each of the determinant categories in the predictive model (cf. Figure 1) were derived from several different sources in the scientific literature which investigated the variables. Moreover, since no single study examined all of the variables simultaneously, only a range of the variance could be estimated for each variable (Kennedy & Smith, 1995). Kennedy, Berbaum, Dunlap, and Smith (1995) also pointed out that there was insufficient information available in the literature on which to identify the interrelationships among the variables, so potential interactions between the variables could not be depicted in their model. Thus, the authors noted that the sum of the variances shown in the model could exceed 100%.

As shown in Figure 1, the potential causative drivers that have been explored to date include those factors related to characteristics of the individual user, length of exposure, usage

schedule, variations in scene content, and features of the equipment (Kennedy, Drexler, & Compton, 1997; Stanney, Kennedy, & Drexler, 1997). Each of these causal drivers of sickness in VEs and their corresponding factors are discussed in the following sections, which follows the Kennedy et al. (1995) taxonomy for organization of the material.

Individual Characteristics

One of the largest contributing factors to VE sickness relates to the characteristics of the individual using the VE. Research has shown that there are large differences in susceptibility to motion sickness where some individuals may be totally incapacitated by exposure to a particular motion stimulus while others remain unaffected (Kennedy, Dunlap, & Fowlkes, 1990; Kennedy, Hettinger, & Lilienthal, 1990). The user characteristics which have been identified as potential factors affecting an individual's susceptibility to provocative motion environments include: age, prior experience, fitness level, gender, and perceptual style (Kennedy, Drexler, & Compton, 1997; Kennedy, Stanney, & Dunlap, 2000; Stanney, Salvendy et al., 1998).

Age

Reports from the scientific literature on motion sickness have indicated that susceptibility to motion sickness fluctuates with age (Reason & Brand, 1975). In general, the findings on age differences in susceptibility have shown: infants (i.e., less than two years old) are virtually immune to motion sickness; children between two to twelve years old are more susceptible to motion sickness than persons 12 to 21 years of age; and thereafter, a gradual decrease in motion sickness susceptibility occurs with increasing age (Kennedy & Frank, 1985; Money, 1970; Reason & Brand, 1975). In contrast, an exploratory field study to reveal factors that may interact

with age in users enjoyment of a VR game was conducted at a VR amusement center (Allen, Singer, McDonald, & Cotton, 2000). The results of the study failed to reveal a significant difference in SSQ sickness scores between three age groups: Young (10-14 yrs.), Middle (21-33 yrs.) and Old (36-36). However, the authors cited several methodological issues to explain the non-significant results including a small sample size, minimal stimulus exposure time (i.e., 5 min. per ride), and a long duration between exposures (i.e., an average of 39 min. between rides).

Experience

Prior experience has also been shown to affect an individual's susceptibility to motion sickness (Kennedy, Berbaum, Allgood, Lane, Lilienthal, & Baltzley, 1988; Pausch, Crea, & Conway, 1992;). Reports from various studies on simulator sickness indicate that more experienced pilots and instructors (i.e., more flight hours in the actual aircraft) had significantly higher incidences of sickness than less experienced (i.e., novice) pilot trainees (Crowley, 1987; Kennedy, Hettinger, & Lilienthal, 1990; McCauley, 1984; Miller & Goodson, 1960). Moreover, Wright (1995) suggested that highly experienced pilots reported simulator sickness at a rate of 150% more than pilots with limited flight experience. Although, Ungs (1988) did not find a significant effect of flight experience on simulator sickness.

Several researchers have postulated that the difference in sickness incidence between experienced and novice pilots is related to the level of familiarity with the actual aircraft (Kennedy, Berbaum, Dunlap, & Smith, 1995; Kennedy, Hettinger, & Lilienthal, 1990; Pausch, Crea, & Conway, 1992). These authors suggested that the more experienced an individual is with the real aircraft, the more apparent any visual or motion discrepancies will be in the simulated vehicle. Relatedly, differences in past experiences with motion sickness have also

been successful in predicting motion sickness incidence (Kennedy, Dunlap, & Fowlkes, 1990). In particular, empirical studies have shown that scores on the Motion History Questionnaire (MHQ), a paper-and-pencil questionnaire used to assess an individual's past history of sickness in various provocative motion environments, are reliable predictors of sickness symptoms in VE systems (Kennedy, Berbaum et al., 1988; Kennedy, Fowlkes, Berbaum, & Lilienthal, 1992; Kennedy, 1996).

Fitness

Because all forms of motion sickness are considered cumulative (i.e., sickness summates), an individual's current physiological state or fitness level can influence their susceptibility to motion sickness (Kennedy, Berbaum, Dunlap, & Smith, 1995; Kennedy, Frank, & McCauley, 1985; Pausch, Crea, & Conway, 1992). Reports from the literature on motion sickness indicate that illnesses such as a cold, flu, or ear infection as well as conditions such as sleep loss, fatigue, or hangover, which are present prior to stimulus exposure, can increase the severity of motion sickness symptoms during or after exposure (DeWit, 1957; Kennedy, Hettinger, & Lilienthal, 1990; Wright, 1995). Accordingly, Kennedy, Lane et al. (1993) have recommended that persons not in their usual state of fitness (i.e., reporting any of the previous conditions) should not be exposed to VE systems (see also Kennedy, Frank, & McCauley, 1985; Kennedy, Hettinger, & Lilienthal, 1990).

Gender

Gender has been implicated as another factor that may influence sickness susceptibility. There is evidence to suggest that women are generally more susceptible to all forms of motion

sickness, including sickness related to VE exposure (Flanagan, May, & Dobie, 2005; Kennedy, Lanham, Massey, Drexler, & Lilienthal, 1995; Kennedy, Stanney, Dunlap, & Jones, 1996; Lentz & Collins, 1977; Reason & Brand, 1975). Hypotheses regarding gender differences in susceptibility have included hormonal influences such as menstruation and pregnancy (Money, 1970; Reason & Brand, 1975), field of view (i.e., women generally have larger fields of view than men; Kennedy & Frank, 1985; Kennedy, Frank, & McCauley, 1985), and perceptual style which is discussed in the next section (i.e., females are typically more field-dependent than males; Kennedy, Lanham et al., 1995). A reporting bias has also been suggested as a potential factor in the gender differences in reported susceptibility. For example, Park and Hu (1999) found that women reported a significantly greater incidence of motion sickness history than did men, but they found no significant gender differences in severity of motion sickness symptoms during exposure to a provocative stimulus (i.e., a rotating optokinetic drum). The authors suggested that the contradictory findings could have been due to social factors (i.e., it is thought to be more socially acceptable for women to admit symptoms of motion sickness than for men), which influenced the differences found in the results on the motion sickness history reports. Dobie, McBride, Dobie, and May (2001) investigated the role of several variables, including exposure history, physical activity, and reporting bias, on gender differences in motion sickness susceptibility. While their results showed that female subjects reported significantly more motion sickness susceptibility, the findings suggested that the differences in susceptibility could not be accounted for by differences in exposure history or physical activity, and there was little evidence to suggest a difference in attitudes of response (i.e., reporting bias) between men and women. In contrast, Graeber and Stanney (2002) suggested that the reported differences in motion sickness between males and females may be due to differences in susceptibility rather

than gender. To test their hypothesis, the authors conducted an empirical study of visually induced motion sickness in an optokinetic (vection) drum that balanced susceptibility level (low versus high) within gender and treatment groups. The results showed there was no significant difference in the severity of sickness between genders, but there was a significant difference in sickness between susceptibility levels.

Perceptual Style

Several empirical investigations have suggested that an individual's perceptual style (i.e., field-dependence/independence) can affect their susceptibility to motion sickness (Kennedy, Drexler, & Compton, 1997; Kennedy, Stanney, & Dunlap, 2000; Stanney, Salvendy et al., 1998). In general, the findings have shown that field-independent individuals were more susceptible to motion sickness than field-dependent individuals (Kennedy, 1975; Kennedy & Frank, 1985; Reason & Brand, 1975).

Exposure Duration

As mentioned previously, scientists generally agree that motion sickness accumulates (Hettinger, Lilienthal, Kennedy, Berbaum, & Hooper, 1987), which suggests that symptoms of sickness will increase as the duration of stimulus exposure increases. In particular, the findings from the scientific literature on motion sickness indicate that as the length of stimulus exposure increases, there is a corresponding increase in the severity of sickness (Kennedy, Berbaum, Dunlap, & Smith, 1995). For example, an analysis of the sickness data collected from pilots exposed to 14 different flight simulators showed a correlation of $r = 0.50$ ($p < 0.05$) between exposure duration and average SSQ Total Severity score (Kennedy & Fowlkes, 1992). Likewise,

Kennedy, Stanney, and Dunlap (2000) reviewed the scientific literature related to the effect of exposure duration on motion sickness and found that, in general, the longer a person is exposed to the sickness stimulus, the greater the incidence of sickness (cf. Ungs, 1988 as an exception). However, the authors indicated that the literature they reviewed contained a limited amount of quantitative data relating exposure duration to motion sickness. Kennedy, Stanney, and Dunlap, therefore, examined the sickness data from a large rotary wing (i.e., helicopter) database which contained approximately 900 cases. The results of their analysis revealed a significant positive linear relationship between exposure duration and simulator sickness (i.e., as exposure duration increased, reported sickness also increased).

Nelson, Bolia, Roe, and Morely (2000) also reported an effect of duration on sickness related to use of a see-through HMD. Their results revealed that the SSQ Total Severity score and scores on the Nausea and Oculomotor SSQ subscales increased as the total time on task increased. Conversely, Stanney and Kennedy (1998) examined the SSQ symptom subscales, which were collected from three different exposure duration groups using an HMD-based VR system, and found a significant effect of exposure duration on sickness, but in the opposite direction. Their results showed that scores on the Disorientation subscale were significantly greater for the 15-minute exposure group than the 30- and 45-minute exposure conditions. Although the results were not statistically significant, the authors also found similar trends across scores on the Nausea and Oculomotor subscales as well as the Total Severity scores. In general, however, exposure duration is considered to be one of the most effective ways to control the severity of sickness because of its cumulative effect on sickness (Stanney, Kennedy, & Drexler, 1997). It should be noted, however, that the length of exposure is often dependent upon the

purpose of the simulation. For instance, training applications typically require longer exposure periods than other applications of VE technology such as research or entertainment.

Usage Schedule

In situations where multiple exposures to a particular VE stimulus are required (e.g., pilot training), the usage schedule relates to the distribution of exposures over a given period time (i.e., the amount of time between exposures). Since the nervous system is relatively adaptive, repetitive stimulation normally reduces the response of the nervous system (Kennedy, Stanney, & Dunlap, 2000). The consequence of this adaptation with respect to motion sickness, including sickness from exposure to VE systems, is that repeated exposures to a provocative stimulus generally reduces the severity of motion sickness in subsequent exposures (Kennedy, Berbaum, Dunlap, & Smith, 1995; Money, 1970; Welch, 2002). For instance, Kennedy, Hettinger, and Lilienthal (1990) reported that individuals with an extensive amount of time in a given flight simulator were less likely to experience symptoms during subsequent exposures to the same simulator.

Kennedy, Stanney, and Dunlap (2000) analyzed SSQ sickness data obtained from pilots exposed to seven consecutive flights in a single helicopter simulator. As predicted, their analysis revealed a significant negative linear trend in sickness as a function of flight number (i.e., as the number of exposures increased, the severity of sickness decreased). Moreover, the authors' review of the literature concerned with the effect of repeated stimulus exposure on various forms of motion sickness suggested an increased tolerance to sickness which occurred for intersession intervals ranging from two to six days (see also Kennedy, Lane et al., 1993). Thus, a desensitization to sickness (i.e., adaptation) can generally be facilitated using short, repeated

exposures to the provocative stimulus that occur close together in time (Kennedy, Stanney, & Dunlap, 2000; Stanney, Kennedy, & Drexler, 1997; Stanney, Kennedy, & Kingdon, 2002).

It should be noted, however, that some researchers (e.g., Kolasinski, 1995) considered repeated exposures to the same provocative stimulus as prior experience and thus, would classify the usage schedule as a "user characteristic" (cf. also Stanney, Salvendy et al., 1998). Nevertheless, the usage schedule is related to the purpose of the VE simulation. For example, a VE that is used for entertainment purposes such as a simulator at a theme park will generally be used only once by a given individual. In contrast, a VE device that is used for training applications, particularly a flight simulator, will be used multiple times by a single individual.

Kinematics

Kinematics refer to the amount of motion in a simulated visual scene which can be affected by factors such as abrupt changes in scene content (e.g., dives, turns), linear/rotational acceleration, or position tracking errors (Kennedy, Berbaum, Dunlap, & Smith, 1995). The perception of illusory (i.e., visually induced) self-motion, known as vection, has been implicated as a primary factor in simulator sickness; the magnitude of vection experienced by an individual predicted the severity of symptoms (Hettinger, Berbaum, Kennedy, Dunlap, & Nolan, 1990). Additionally, previous research by Hettinger, Owen, and Warren has shown that variations in optical flow rate and texture density (i.e., scene detail) affected the strength of vection (as cited in Kennedy, Berbaum et al., 1995 and Kennedy & Smith, 1995).

Several investigations were conduced by Kennedy and his colleagues in an effort to identify and quantify the type of visual motion stimulation, especially linear and rotational velocity cues, that affected simulator sickness (Kennedy, Berbaum, Dunlap, & Hettinger, 1996;

Kennedy, Berbaum et al., 1995; Kennedy, Berbaum, & Smith, 1993; Kennedy & Smith, 1995). The specific objectives of their research were to identify the parameters involved in visually specified motion (i.e., the kinematic elements) and then establish a relationship between the visual motion parameters and sickness. The authors predicted that the magnitude of kinematics would affect sickness in a VE; more dynamic visual motion would increase symptom severity.

In order to quantify motions within the visual scene, more than a dozen different kinematic variables related to scene complexity, depth and distance cues, and the amount of visually presented roll (tilt), yaw (turn), and pitch (ascend/descend) were collected from the visual display during flight scenarios in three different simulators with markedly different visual systems (Kennedy, Berbaum et al., 1995). Data for the kinematic variables and simulator sickness symptoms were then analyzed in order to determine the provocativeness of the different variables with respect to simulator sickness. The results from the investigations indicated that visual kinematics (e.g., edge rate, roll rate, etc.) could be related to the severity of simulator sickness (r = 0.30 to 0.40), but the automated scoring method used for analyzing the kinematic data was only able to identify a clear relationship when the stimulus was non-interactive (Kennedy, Berbaum et al., 1995). Because simulator and VR exposures generally provide a very interactive visual environment (i.e., the user controls part of the visual stimulus), the researchers concluded that additional research was needed in order to develop kinematic measures which could take into account the interactive nature of VE stimulus conditions.

Equipment Features

Specification of the equipment parameters that promote effective performance and realism, but avoid or minimize sickness is critical for the design and use of VE systems

(Kennedy, Berbaum, & Smith, 1993). A number of design inadequacies or equipment limitations have been reported in the scientific literature as potential factors which contribute to sickness in VEs. In the following sections, the equipment features implicated as factors influencing sickness are presented and categorized according to the type of VE system in which they can be found. The features which are common to VR and simulator systems are presented first, followed by the features which are specific to VR systems, and the features which are specific to simulators. It is important to note that although see-through HMDs (designed for augmented reality applications) and desk-top displays can be classified as VE systems, they are beyond the scope of this research, which is focused on more immersive-based VE systems, and thus were not included in the subsequent review.

Features Common to VR Devices and Simulators

Individuals largely rely on their visual senses during exposure to a VE system and as such, the visual display will provide the most salient and detailed information about the simulated environment (Durlach & Mavor, 1995; Wilson, 1997). The visual display not only provides 'input' to the user, changes in the visual scene also represent the 'output' of the user (Kennedy & Smith, 1996). However, Ebenholtz (1992) stated that VEs are very interactive and as a result, the visual display system engages "numerous oculomotor systems, and hence have the potential to produce motion sickness symptoms" (p. 303). The characteristics related to image presentation in VE systems that have been implicated as factors influencing sickness include the field-of-view, display resolution, viewing region, and different types of temporal delays (refresh rate, update rate, and system latency).

Field-of-View

Field-of-view (FOV) refers to the horizontal and vertical angular dimensions of a visual display (Pausch, Crea, & Conway, 1992). Research has shown that wider FOVs provide better task performance (Bowman, Datey, Ryu, Farooq, & Vasnaik, 2002; Pausch, Crea, & Conway, 1992; Wilson, 1997). However, the size of the FOV has been implicated as a critical causal factor in simulator sickness (McCauley, 1984). In general, research on the effects of sickness related to FOV size indicate that wider FOV displays increase the incidence and intensity of simulator sickness, particularly symptoms of eyestrain, headache, and dizziness (DiZio & Lackner, 1992; Hettinger et al., 1990; Lawson et al., 2002; Padmos & Milders, 1992; Pausch, Crea, & Conway, 1992; Rinalducci, 1996). For example, Lin, Duh, Parker, Abi-Rached, and Furness (2002) conducted an empirical study to examine the effects on sickness as a function of varying the display FOV. Participants wore CrystalEyes stereo glasses and were exposed to four different FOVs (60°, 100°, 140°, and 180°) in a driving simulator. Their results revealed that sickness severity increased with increasing FOVs, although the scores at 140° and 180° were not significantly different.

Bowman, Datey, Ryu, Farooq, and Vasnaik (2002) asserted that FOV is usually a trade-off with resolution. Wider FOVs can produce poor visual resolution because the available pixels are more spread out (Bowman et al., 2002; Wilson, 1997). In contrast, narrower FOVs (i.e., 40-60° vertical by 60-80° horizontal) with higher resolution can cause 'tunnel vision' or increase disorientation effects (Bowman et al., 2002; Wilson, 1996). Relatedly, Kennedy, Fowlkes, and Hettinger (1989) indicated that wide FOV displays can magnify the effects of any distortions in the visual display. Durlach and Mavor (1995) also noted that greater geometric image distortions occur in HMD displays with large fields of view because a greater degree of

magnification is required to project the real-world size image onto the small display screens. Other research related to FOV size, discussed below, has suggested that the incidence of sickness is influenced by the amount of vection or flicker produced by the display.

Vection

The research literature from various types of vection studies, including those involving exposure to VE systems, has shown that motion sickness is a common side effect of viewing visual scenes of self-motion without actual physical movement (Hettinger, 2002; Hettinger et al., 1990; Hettinger & Riccio, 1992; Yardley, 1992). As mentioned previously, vection is the illusion of self-motion in the absence of actual physical movement which can induce symptoms of motion sickness (Hettinger et al., 1990). However, while vection has been correlated with visually induced sickness, Lawson et al. (2002) maintained that vection is not a necessary precursor of symptoms. Specifically, not all people who experience vection will experience motion sickness, but those who do experience vection are more likely to experience sickness (Hettinger & Riccio, 1992; Hettinger et al., 1990).

Several researchers have reported that displays with a wide FOV provide a more compelling sensation of vection as well as a better orientation within the simulated environment (Hettinger et al., 1987, 1990; Kennedy, Fowlkes, & Hettinger, 1989; Padmos & Milders, 1992; Pausch, Crea, & Conway, 1992). Yardley (1992) also suggested that as the area and velocity of motion in the visual field increased, there would be a corresponding increase in the experience of vection. Kennedy, Hettinger, and Lilienthal (1990) indicated that peripheral vision is particularly sensitive to motion stimulation. Therefore, wider fields of view enhance the experience of vection because stimulation of peripheral vision is more effective in inducing self-

motion than stimulation of central vision (Hettinger, 2002). Because a wide field of view is more conductive to inducing vection, it is also more likely to produce motion sickness symptoms (Hettinger et al., 1990; Hettinger & Riccio, 1992). Moreover, Durlach and Mavor (1995) reported that greater levels of motion sickness are produced when users make head movements in VE displays that induce vection. Although the majority of the scientific literature indicates that a wide FOV can induce motion sickness, Hettinger et al. (1987, 1990) cited a study by Andersen and Braunstein where reports of vection and motion sickness were found using a display with a relatively small FOV (i.e., a 7.5° visual angle).

Flicker

Durlach and Mavor (1995) stated that sensitivity to flicker is greater in peripheral vision than in foveal (i.e., central) vision (see also Boff & Lincoln, 1988c). Thus, the size of the FOV can also affect flicker perception. In particular, a wider FOV display will increase the likelihood that the user will perceive flicker because more of the peripheral vision will be stimulated (Durlach & Mavor, 1995; La Viola, 2000; Pausch, Crea, & Conway, 1992). Flicker is not only distracting to the VE user, it can also induce symptoms of motion sickness, particularly those related to the visual system (La Viola, 2000).

Display Resolution

Pausch, Crea, and Conway (1992) defined the resolution of a visual display as the amount of detail provided by the display (i.e., the image quality) which is measured in pixels per inch. In HMD VR systems, the most frequently used type of display screen is a back-lighted liquid crystal display (LCD; Durlach & Mavor, 1995). Wilson (1996) reported that in many HMDs the

number of pixels per LCD will range from 360 x 240 to 720 x 480 which can support stereoscopic vision. However, Piantanida et al. (as cited in Wilson, 1996) equated these spatial resolution values to a visual acuity ranging from 20/200 to 20/100 which can affect object perception. Moreover, poor resolution can cause strain on the visual system as the user tries to focus on the simulated images and poor resolution can therefore produce symptoms such as eyestrain and headache. As mentioned above, resolution is usually a trade-off with FOV because in wider FOV displays the available pixels are more spread out over the retinal area stimulated which reduces display resolution (Bowman et al., 2002). Accordingly, in simulators with computer-generated image (CGI) display systems which have a fixed pixel capacity, high spatial resolution may be limited to a small FOV (Rinalducci, 1996).

Viewing Region

The viewing region of a display is the area in which the system user is able to maintain an image of the simulated scene (Padmos & Milders, 1992). The design eye point, also referred to as the design eye, is the point located in the center of the viewing region which is the optimal position for the user to view the display (Pausch, Crea, & Conway, 1992). Kennedy, Fowlkes, and Hettinger (1989) explained that graphic displays such as those used in simulator visual systems only provide an accurate visual representation when they are viewed from the geometric center of the projection (i.e., the design eye; cf. also Kennedy, Berbaum et al., 1987). Consequently, the visual image becomes increasingly distorted as the eccentric distance from the design eye point increases (Padmos & Milders, 1992; Pausch, Crea, & Conway, 1992). For example, Kennedy, Fowlkes, and Hettinger (1989) cited previous findings where pilots that

viewed the flight simulator display from an off-axis position (i.e., alongside and eccentric to the design eye) experienced more simulator sickness as a result of viewing distorted images.

It is also possible for more than one person to be simultaneously exposed to a simulator, particularly flight simulators (e.g., co-pilot, flight engineer, or flight instructor). However, Kennedy, Berbaum et al. (1987) noted that the design eye in simulators is typically about a one-quarter cubic foot of space. Thus, while these other participants may be within the viewing region of the display, they could positioned outside of the design eye point which would increase the likelihood of simulator sickness (Pausch, Crea, & Conway, 1992). Additionally, Kennedy, Fowlkes, and Hettinger (1989) reported that detailed visual imagery and wide field-of view displays can magnify the visual distortion caused by viewing the display from outside of the design eye point.

Piantanida (as cited in Wilson, 1996) indicated that optical distortion can also occur in VR systems when there is a discrepancy between the interpupillary distance (IPD) of the user and the optical centers of the HMD display screens (cf. also Mon-Williams, Wann, & Rushton, 1993; IPD is discussed further in a later section). Moreover, Piantanida suggested that optical distortions are generally likely with HMD-based systems because the lenses are imperfect. Similar to the design eye in simulators, prismatic distortions from the lenses could occur if the individual is not looking through the center of the lenses such as when the headset is not properly adjusted or while the participant looks around the visual environment (Wilson, 1996). Accordingly, Durlach and Mavor (1995) declared that the optics in an HMD must allow for clear focusing and off-axis viewing.

Relatedly, a high degree of optical magnification is required to transfer the simulated scene on the small display screens within the HMD into a real-world size image on the retina

(Durlach & Mavor, 1995). Moreover, because the displays screens are positioned about an inch in front of the eyes (i.e., a fixed close viewing distance), greater geometric image distortions occur as the degree of magnification increases (Durlach & Mavor, 1995).

Temporal Delays

Simulators and VR devices are controlled by computer systems which must perform a large number of calculations in order to generate the simulated visual imagery, control the inertial or position tracking system, as well as to monitor and respond to the control inputs of the system user (Frank, Casali, & Wierwille, 1988). Therefore, as the number of required calculations increase, the temporal delay between an operator's input to the system and subsequent changes in the system output, in terms of the visual display and motion base, can also increase. For example, an increase in scene complexity requires more calculations by the computer and thus, can increase temporal delay (Frank et al., 1988). Other factors that can affect computational and rendering speeds include wider FOV displays, higher image resolution, and visual scene changes which accommodate head movements (Durlach & Mavor, 1995). Moreover, Frank et al. (1988) asserted that separate computers with different update rates are often used for the visual and motion systems in simulators which can exacerbate temporal delays and thereby make the visual-inertial delays asynchronous.

While temporal delays can obviously affect the performance of the system user, temporal lags in VE systems also have the potential to contribute to motion sickness (Wilson, 1997). Wilson (1996) also suggested that faster VE systems could actually cause more problems than slower systems depending on the temporal lags present in the system. The factors that limit

temporal resolution include display refresh rate, update rate, and system latency (Durlach & Mavor, 1995).

Refresh Rate

Refresh rate, or frame rate, is defined as the frequency with which an image is generated on the display, that is the time required to update the visual image on the screen (Blade & Padgett, 2002; Padmos & Milders, 1992; Pausch, Crea, & Conway, 1992; Wilson, 1996). Durlach and Mavor (1995) asserted that the interactive nature of VEs require high frame rates. In general, they indicated that the specific frame rate required in any particular situation depends on the type of environment simulated. For example, the authors stated that the frame rate for relatively static environments with slow moving objects should not be less than 10 frames per second with a total system latency not more than a tenth of a second (i.e., 100 msec). In contrast, environments that include objects with relatively high frequencies of motion will require significantly higher frame rates (i.e., greater than 60 Hz) and much shorter system delays (e.g., 17 msec).

The refresh rate can affect the quality of the displayed images, but is also related to the perception of flicker (Durlach & Mavor, 1995; Wilson, 1996). Specifically, the refresh rate can interact with luminance (i.e., the brightness or intensity of the light coming from the display) to produce flicker which contributes to visual fatigue and simulator sickness (Padmos & Milders, 1992; Pausch, Crea, & Conway, 1992). For instance, higher luminance levels and higher contrast levels are known to increase flicker sensitivity while slower refresh rates can promote flicker in the visual display (Boff & Lincoln, 1988b; Durlach & Mavor, 1995; Padmos & Milders, 1992; Pausch, Crea, & Conway, 1992). Therefore, the refresh rate must be high enough

to avoid flicker (Kennedy, Berbaum, Dunlap, & Smith, 1995). However, because of the interaction of refresh rate, luminance, and contrast, in order to suppress flicker the refresh rate must increase as luminance and contrast increase or vice versa (Pausch, Crea, & Conway, 1992).

Durlach and Mavor (1995) asserted that the typical luminance level in HMD displays were sufficient to cause flicker for frame rates of 30 Hz or less. Boff and Lincoln (1988a) also noted that sensitivity to flicker is greatest for frequencies between 5 - 20 Hz. Moreover, Boff and Lincoln (1988d) indicated that displays with refresh rates less than 20 Hz can create flicker that is usually "quite annoying to the observer" (p. 2258), and disorientation and confusion may occur with refresh rates between 7 -15 Hz. Accordingly, Boff and Lincoln (1988d) noted that for most electronic displays, flicker perception could be eliminated in the fovea if the refresh rate is 35 Hz or higher whereas a frame rate of at least 47 Hz was required to eliminate flicker for peripheral viewing. Similarly, La Viola (2000) suggested that perceived flicker could be eliminated in the fovea with a 30 Hz refresh rate, but a higher refresh rate was required to eliminate flicker in the periphery for large targets. Since sensitivity to flicker increases with larger fields of view, faster refresh rates (i.e., 80-90 Hz) may also be required in field sizes larger than 70° in order to avoid flicker (Padmos & Milders, 1992). Therefore, May and Badcock (2002) suggested that with current display luminances, a frame rate of at least 120 Hz was required to avoid flicker (see also Bridgeman, 1995).

Update Rate

Update rate is defined as the rate or frequency with which a new image is generated and shown on the visual display and is typically measured in frames per second (fps; Durlach & Mavor, 1995; Padmos & Milders, 1992; Pausch, Crea, & Conway, 1992; Wilson, 1996). The

update rate is determined by the power of the computer hardware (i.e., the computational speed) and is inversely related to the complexity of the visual scene (Dulach & Mavor, 1995; Pausch et al., 1992; Wilson, 1996). In other words, there is a trade-off between screen update rate and visual scene complexity where faster update rates limit the level of visual complexity available (Padmos & Milders, 1992; Wilson, 1997). For example, Wilson noted that a 30 fps update rate is a 'comfortable' rate for the eye because it is similar to a watching a video, but more detailed and complex applications can only support 10-20 fps.

A low update rate can cause the images in the visual display to shake and create contour distortions (Padmos & Milders, 1992). Furthermore, inadequate display update rates can produce disorientation and other symptoms of motion sickness (May & Badcock, 22002). For example, Durlach and Mavor (1995) indicated that update rates below 12 Hz can induce motion sickness. Therefore, the minimum update rate that has been proposed for use in VR systems is 12 fps in order for the display motion to be perceived as smooth and to provide some realism in the visual dynamics (Durlach & Mavor, 1995; Wilson, 1996). Although, Durlach and Mavor maintained that the ideal update rate is 20 fps or higher.

In CGI simulator displays (discussed in a later section), the maximum update frequency also depends on the complexity of the visual scene (i.e., the number of polygons to be processed) as well as the total number of pixels that can be processed each second (i.e., the pixel fill rate; Padmos & Milders, 1992). The authors noted that 30 Hz would be a sufficient update frequency for many simulator applications, but higher update frequencies would be required when faster angular speeds of displayed objects were used in order to avoid shaking images. However, Wilson (1996) indicated that update rate and system latency (discussed in the next section) are

independent, so even with a fast update rate there may still be lags in the system which can cause disorientation.

System Latency

Simulators and VR devices are computer-based systems, so computational limitations of the equipment can produce a temporal delay between operator input and subsequent changes to the visual display (Kennedy, Fowlkes, & Hettinger, 1989). In the scientific literature, various terms have been used for this type of delay including system lag/latency, system update rate, image delay, or transport delay. Padmos and Milders (1992) noted that system latency is a combination of: (1) the sampling time of the operator input controls, (2) the time to calculate a viewpoint change, and (3) the time between position change input from the host computer to the visual display system and rendering the corresponding image.

A large degree of system latency can affect the operator's control of the simulated environment and it can increase simulator sickness (Padmos & Milders, 1992). Previous research in flight simulators has shown that when large system delays were present, the pilot was unable to accurately predict the length of the delay which caused the pilot to base their current actions on a guess of the vehicle's position as a result of their previous control input (Pausch, Crea, & Conway, 1992). The authors reported that this technique, sometimes referred to as "guess and lead the system", usually failed and caused the pilot to overcompensate control of the vehicle which produced oscillations. Consequently, abnormal accelerations caused by the operator-induced oscillations increased the potential for simulator sickness because very low frequency motion or visual distortions were produced as a result of the increased load on the computer running the simulator (Kennedy, Berbaum, Dulap, & Smith, 1995). Accordingly,

Padmos and Milders (1992) recommended that system delays should be no more than 40-80 ms in driving simulators and 100-150 ms in flight simulators.

In VR systems, system lag or latency is defined as the amount of time needed to send a signal from the position tracker (discussed further in the next section) and subsequent presentation of the image on the display screen (Wilson, 1996). In other words, the time between when an individual moves within the environment and when the movement is reflected in the visual scene. Accordingly, system lag in VR systems is composed of the position tracker delay, the delay in sending the position information to the computer, and the delay in processing the information and creating the image (Wilson, 1996). However, Pimentel and Teixeira (as cited in Wilson, 1996) reported that system latencies of 100 ms or greater caused motion sickness symptoms.

DiZio and Lackner (1997) investigated the effects of system delay (i.e., delay between head movements and updates to the visual scene) on motion sickness. Participants were exposed to a *stationary* visual scene in an HMD and asked to make paced voluntary head and eye movements in order to view a series of landmarks. The experimental conditions varied system update delay (67, 100, 200, and 300 ms) and field-of-view (wide [126° x 72°] versus halving the linear dimension). The study found that significant motion sickness symptoms, including nausea, were induced in the shortest delay condition and the severity of sickness increased monotonically with system delay. However, the results also showed that reducing the field-of-view reduced the effect of the update delay on sickness. That is, the severity of motion sickness was cut in half in the decreased field-of-view condition with a 200 ms system delay. While the Dizio and Lackner study provided important insights on the relationship between system delay and field-of-view, there is a caveat with respect to their findings. The study methodology (i.e., a

within-subject design, relatively small sample size [n = 20], and only one HMD with very specific system parameters), may prohibit replication of the findings in a subsequent study as well as limiting the generalizability of the results to other types of VE systems.

Features Specific to VR Devices

The equipment features that are specific to VR devices and which have been implicated as factors influencing sickness include the type of display (binocular and bi-ocular), interpupillary distance, helmet weight, and the position tracker (errors and latency).

Type of Display

Helmet-mounted displays (HMDs) typically contain two liquid crystal displays (LCDs) with magnifying optics which are positioned in front of each eye (Rinalducci, 1996). The displays are either stereoscopic binocular displays or bi-ocular displays (Mon-Williams & Wann, 1998; Wann & Mon-Williams, 2002). Binocular displays present a slightly different image to each eye with some degree of overlap (about 60°) which provides stereoscopic depth information (i.e., cues for the distance of objects) similar to viewing objects in the real world (Mon-Williams & Wann, 1998; Rinalducci, 1996; Wilson, 1996). Conversely, bi-ocular displays present identical images to each eye so the depth cues that are available from stereoscopic displays are not provided in biocular systems (Mon-Williams & Wann, 1998; Pausch, Crea, & Conway, 1992; Rinalducci, 1996).

Because humans have two eyes with some degree of spacing between them, under normal viewing conditions, a slightly different image is seen by the two eyes when viewing an object which is called retinal disparity (Rinalducci, 1996). Retinal disparity provides stereopsis, which

is the ability to judge relative depth (i.e., to see very small differences in depth; Rinalducci, 1996). Thus, when viewing a near object, our eyes turn inward together (i.e., convergence) in order to see the object as a single entity and the curvature of the lens changes to focus the image on the retina, which is called accommodation (Ebenholtz, 2001; May & Badcock, 2002; Wilson, 1996). Furthermore, accommodation and convergence are cross-linked so the eyes normally converge and accommodate for the same distance and accommodation produces convergence and vice versa (Mon-Williams & Wann, 1998).

However, in a stereoscopic HMD display, the screens are only positioned about an inch away from the eyes whereas the images presented on the screens can show objects positioned at different optical distances (e.g., 10 ft., 100 ft., etc.; Wilson, 1996). As a result, accommodation is fixed to the distance of the display screen in order to focus the screen images, but the degree of convergence changes relative to the distance of the virtual objects being viewed (Rinalducci, 1996; Wann & Mon-Williams, 2002). Therefore, the normal accommodation-convergence relationship is disrupted because there is a mismatch between the amount convergence and accommodation need to view the display which can cause symptoms such as eyestrain or headache (Ebenholtz, 1992; Kennedy, Berbaum et al., 1987; Wilson, 1996). Durlach and Mavor (1995) also asserted that the relation between convergence and accommodation can influence the distortion of images.

Several empirical studies have evaluated the effects of binocular and bi-ocular system use on the visual system. Mon-Williams, Wann, and Rushton (1993) examined the effects of using a binocular (stereoscopic) HMD on the visual system. The results of various ophthalmic tests of binocular function revealed deficits in binocular vision after a relatively brief exposure (i.e., 10 minutes) to the HMD (cf. Wann & Mon-Williams, 2002 for a detailed description of the tests).

Participants also reported symptoms related to disturbances of the visual system including blurred vision, eyestrain, headache, and difficulty focusing. Several of the participants also reported experiencing motion sickness, especially nausea.

Rushton, Mon-Williams, and Wann (1994) hypothesized that the primary cause of the visual deficits found in the Mon-Williams et al. (1993) study was the conflict between the stereoscopic depth cues, image disparity and focal depth (i.e., the information that produced a conflict in accommodation and convergence). Therefore, Rushton et al. replicated the Mon-Williams et al. study using a bi-ocular display and a larger sample size. Bi-ocular displays present the same image to each eye, so there is no dissociation between convergence and accommodation (Wilson, 1996). In contrast to the Mon-Williams et al. study, no significant changes in visual performance were found on the battery of ophthalmic tests for exposure periods of up to 30 minutes. Additionally, compared to the motion sickness symptoms found in the previous study, mild symptoms of visual strain were only reported by a few participants. Although the bi-ocular HMD system differed from the binocular system (e.g., IPD adjustments, independent eye focus, higher screen resolution, and less temporal lags), the authors believed that a crucial difference in the visual effects was due to the difference in the type of display (i.e., bi-ocular versus binocular).

Mon-Williams and Wann (1998) later demonstrated that even during relatively short exposures (i.e., 10 minutes) to a binocular HMD display, a continual conflict between accommodation and convergence caused stress on the visual system. Study participants reported adverse visual symptoms (e.g., eyestrain, headache) and measurable changes in visual functioning were found on the battery of ophthalmic tests. Therefore, Mon-Williams and Wann (1998) concluded that the differences in effects on the visual system between binocular and bi-

ocular displays found in their previous studies was due to accommodation-convergence conflicts rather than the stereoscopic depth information provided in binocular displays. Based on their findings, the investigators also expressed concern that the changes they found in participants' visual functioning due to exposure to the HMD could affect subsequent performance on visually demanding tasks such as driving. Thus, stereoscopic systems may support better task performance, but they also increase the likelihood for visual side effects compared to bi-ocular displays because of the inherent conflict between accommodation and convergence (Wann & Mon-Williams, 2002Wilson, 1996).

Interpupillary Distance

Some HMDs provide the ability to adjust the lateral distance between the eyepieces (i.e., the display screens) in order to accommodate differences in the interpupillary distance (IPD) of the users, but others only provide a fixed distance between the optical centers of the display lenses (Mon-Williams, Wann, & Rushton, 1993). However, as mentioned previously, a discrepancy between the IPD and the optical centers of the display screens can create optical distortions in the visual imagery (Wilson, 1996). Based on the findings from their research, Mon-Williams and his colleagues declared that an incorrect IPD can induce prismatic visual effects caused by viewing the image off-center which produces stress on the visual system (Mon-Williams, Wann, & Rushton, 1993, 1995; Rushton, Mon-Williams, & Wann, 1994).

Helmet Weight

The weight of an HMD can vary from four ounces to more than five pounds (McCauley-Bell, 2002). However, changing the weight of the head, which alters the inertia of the head, can

be extremely provocative (Durlach & Mavor, 1995). DiZio and Lackner (1992) argued that the weight of an HMD creates sensorimotor rearrangements during head movements which can contribute to motion sickness. They also noted that an HMD which weighs 2.5 pounds increases the effective weight of the head by at least 20%. Similarly, Durlach and Mavor (1995) pointed out that wearing an HMD which increased the weight of the head by 50% can, in general, increase a person's susceptibility to motion sickness during exposure to angular acceleration. For instance, DiZio and Lackner (1992) discussed the results of a study where participants were exposed to periodic angular accelerations and decelerations in a rotating chair. Motion sickness symptoms were more severe in participants wearing a weighted helmet during exposure than those with no load on their head.

Most HMDs are also coupled with a position tracking device which necessitates head movements in order to change the viewpoint of the simulated visual scene. However, Durlach and Mavor (1995) indicated that susceptibility to motion sickness is further increased if voluntary head movements are made while the weight of the head is altered because it makes the movements more provocative. Consequently, the authors declared that "simply wearing an HMD can be provocative in itself, regardless of the scenes displayed" (p. 208). Dizio and Lacker similarly remarked that their observations of participants suggested simply moving around while wearing the HMD elicited some motion sickness symptoms.

Position Tracker

An important component of VR systems is the ability to detect and track the position and orientation of the user's head in order to identify where the individual is looking within the environment so that the appropriate changes can be made to the simulated scene (Durlach &

Mavor, 1995; Wilson, 1996). The majority of HMD visual display systems are directly coupled to the motion of the user's head using a position tracking system (Durlach & Mavor, 1995). A position tracker, consisting of sensors mounted to the HMD, first determines the position and orientation of the user's head and then transfers the information to the processing computer which generates and renders an image that corresponds to a viewpoint change in the simulated scene based on the user's head movements (Biocca, 1992; Wilson, 1996).

Errors

The accuracy of the position information provided by a head tracker can vary, and as a result, the level of inaccuracy in a given tracker can influence the incidence of sickness symptoms (La Viola, 2000). For instance, a study by Bolas (as cited in Wilson, 1996) indicated that nausea was a consequence of "poorly tracked systems, with slow response and noise in the tracking system" (p. 43). Additionally, the stability of the information provided by some tracking devices can produce jitter and thus, distortion in the visual image which can induce symptoms of motion sickness such as disorientation (La Viola, 2000).

Lag

Another temporal constraint of many VE systems is the lag associated with position tracking systems. However, the overall performance of an HMD system (i.e., update rate and lag) is linked to the performance of the position tracking system (Durlach & Mavor, 1995). In fact, delays from position tracker systems were cited as the major factor contributing to update delays in HMD images (Durlach & Mavor, 1995). Moreover, DiZio and Lackner (1992) asserted that temporal distortions in the visual display occur because "the visual displays and

head tracking devices do not match human capabilities and graphics systems cannot keep up with rapid human movements" (p. 322).

The latency of a position tracker is based on the time required to register the user's position or movement and the time to send the information to the processor (Wilson, 1997). Once the signal is received by the processor, there is another delay in processing the position information and rendering the update in the visual scene (Wilson, 1997). However, if position tracker delays are present, the user may perceive a difference in what is represented within the visual scene and what they are doing in the real-world (i.e., a mismatch between head motion and the visual display) which can affect task performance as well as induce symptoms of simulator sickness including nausea or dizziness (Allison, Harris, Jenkin, Jasiobedzka, & Zacher, 2001). For instance, Hettinger and Riccio (1992) indicated that symptoms of motion sickness often occur when detectable and excessive lags are present while using an HMD. Moreover, the position tracker delay can be especially nauseogenic in wide FOV displays because larger head movements are needed to acquire targets in the peripheral field (Durlach & Mavor, 1995). A study by Draper et al. (2001), however, provided an exception to the general findings reported in the literature. In their experiment, two time delays (125 ms and 250 ms) were created using a delay buffer between the head tracker and image processing computer and the effect of the delays on sickness were evaluated. Their findings revealed that sickness symptoms were induced by exposure to the HMD system, but contrary to the investigators' hypothesis, there was no significant effect of time delay on sickness.

Delays between a tracker system acquiring position information and the viewpoint update on the screen can range from 10-250 ms for the electromagnetic tracking systems which are commonly used (Draper, Viirre, Furness, & Gawron, 2001; Durlach & Mavor, 1995). Durlach

and Mavor maintained that tracker-to-host computer rates must be at least 30 Hz because delays between head motion and visual feedback less than 60 ms may induce simulator sickness. Therefore, they argued that position trackers should not contribute more than 10 ms to overall system latency.

Features Specific to Simulators

The equipment features that have been implicated as factors influencing sickness which are specific to simulator systems include CGI displays, collimation, simulator platform, motion frequency, and temporal lag.

CGI Displays

Many flight and driving simulators employ multiple CRT visual displays using computer-generated imagery (CGI; Kennedy, Fowlkes, & Hettinger, 1989). However, misalignment of the CGI optical channels can cause distortion in visual images because the design eye from which all CGI channels could be viewed simultaneously is eliminated (Kennedy, Berbaum et al., 1987 Kennedy & Fowlkes, 1992). Therefore, the same optical distortions that occur when the system operators move their heads outside of the design eye (cf., viewing region section) can be created. Additionally, if the focus of the CGI channels are different, different accommodative distances would be required to view a scene that was imaged at infinity (Kennedy, Berbaum et al., 1987). The authors declared that the consequence of these repeated changes in accommodation can be eyestrain or headache. They also noted that the incidence and severity of eyestrain was higher in flight simulators with CGI displays than in those with dome displays. Moreover, Kennedy,

Berbaum et al. (1987) argued that the number of CGI optical channels was generally proportional to the number of symptoms reported.

Collimation

Collimation relates to the parallel alignment of the light rays emitted by the visual display, which places the image at optical infinity (Padmos & Milders, 1992). In simulators, collimated images from more than one image channel (i.e., display) are often seamlessly combined using concave mirrors (Padmos & Milders, 1992). Collimated images are typically used to increase realism in the simulated environment by creating an illusion of depth in two-dimensional images. Kennedy (1996) explained that an improperly collimated system can produce negative convergence and accommodation which can contribute to simulator sickness, especially symptoms associated with disturbances of the visual system (e.g., eyestrain, headache, etc.). Collimated images used in driving simulators can also negatively affect the distance perception of near objects and create a false perception that the eyes are positioned at an exaggerated height (Padmos & Milders, 1992; Ebenholtz, 1988).

Simulator Platform

The platform of a simulator is either a fixed-base or motion-base. In a fixed-base simulator, information regarding self-motion is provided solely by the visual display system. In contrast, a motion-base simulator provides a subset of the inertial forces that would be present during real movement in the vehicle being simulated (DiZio & Lackner, 1992; Durlach & Mavor, 1995). Specifically, a motion-base simulator can provide two types of inertial cues: acceleration and tilt (Kennedy, Berbaum et al. (1987). McCauley and Sharkey (1992) indicated

that the hydraulic motion-base typically used on simulators provide six axes of movement with ±35° of angular displacement and two meters of linear displacement. Motion-base systems are extremely expensive, but they are used in specific applications (e.g., flight simulators) to enhance the sense of self-motion provided by the visual display (Durlach & Mavor, 1995). However, a motion-base simulator can provide motion cues compatible with initial but not sustained acceleration (Dulach & Mavor, 1995). For example, forward acceleration can be simulated by pitching the base backward while also translating it forward slightly (Durlach & Mavor, 1995).

Visual movement through a simulated environment that is not accompanied by the normal inertial cues (i.e., forces and accelerations) associated with movement through the real environment can induce motion sickness, particularly nausea (Durlach & Mavor, 1995; May & Badcock, 2002; McCauley & Sharkey, 1992). Consequently, the overall incidence of simulator sickness is typically lower in simulators with a motion-base than those with a fixed-base (McCauley, 1984). Kennedy, Berbaum et al., (1987) suggested that one of the reasons simulator sickness incidence was lower in simulators with a motion base compared to fixed-base simulators was because of differences in pilot head movements during exposure. The authors explained that in a moving-base simulator, pilots' head movements were similar to those in the actual vehicle whereas the head movements in fixed-base simulators were often in conflict with the inertial stimulus, which increased the provocativeness of the simulation. There have, however, been a few reports that contradict the general findings of a difference in sickness incidence between fixed-base and motion-base simulators. For example, a study by Sharkey and McCauley (as cited in McCauley & Sharkey, 1992) found a relatively equivalent incidence of simulator sickness in a motion-base helicopter simulator as in the fixed-base simulator.

Motion Frequency

A strong relationship between motion sickness incidence and exposure to very low frequency whole-body vibration has been found in a variety of provocative motion environments including ships at sea, planes, spacecraft, automobiles, buses, trains, and motion-base simulators (Guignard & McCauley, 1990). Research has indicated that the most nauseogenic frequency of motion is centered around 0.2 Hz; the lower limit for nauseogenic motion is frequencies below 0.1 Hz and a decline in acceleration-induced motion sickness also occurs at frequencies above 0.2 Hz (Guignard & McCauley, 1990).

It is generally agreed that simulator sickness incidence in moving-base simulators depends on the frequency and acceleration characteristics of the motion produced by the simulator platform (Kennedy, Hettinger, & Lilienthal, 1990). Specifically, the incidence and severity of sickness is usually greatest when the energy spectra from the motion base is in the nauseogenic very low frequency range of 0.2 Hz (Kennedy, Hettinger, & Lilienthal, 1990; Lawson et al., 2002; McCauley, 1984). Kennedy, Berbaum et al. (1987) also reported that motion sickness is proportional to the acceleration in a system, so 0.2 Hz is more nauseogenic than 0.5 Hz. Moreover, an examination of the sickness rates in several motion-based flight simulators indicated that the simulators which produced linear oscillations in the range of 0.2 Hz (i.e., very low frequency motion) showed significantly higher incidence and severity of simulator sickness than motion-base simulators which had low levels of energy in the 0.2 Hz region (Kennedy, Allgood, Van Hoy, & Lilienthal, 1987; Van Hoy, Allgood, Lilienthal, Kennedy, & Hooper, 1987). Thus, motion-base simulators with acceleration frequencies in the 0.2 Hz range (i.e., very low frequency motion) can be considered a major contributor to simulator sickness (Kennedy, Berbaum et al., 1987).

Temporal Lag

As mentioned previously, simulators are computer-based systems and as such, computational limitations of the equipment can produce temporal lags between operator input and subsequent changes to the visual display, the motion base, or both (Kennedy, Fowlkes, & Hettinger, 1989; McCauley, 1984). Inaccuracies in motion cueing which are created by temporal delays between the control inputs of the operator and subsequent changes in the visual display and motion base have been implicated as a contributing factor to the incidence of simulator sickness (Kennedy & Fowlkes, 1992; Kennedy, Hettinger, & Lilienthal, 1990; McCauley, 1984).

Uliano and his colleagues (Uliano, Kennedy, & Lambert, 1986; Uliano, Lambert, Kennedy, & Sheppard, 1986) evaluated the effect of lag between a pilot's control input and the resulting change to the visual scene on performance and simulator sickness. Three separate visual delays (126 ± 17 ms, 177 ± 23, and 215 ± 70) were presented to pilots in a fixed-base flight simulator with a wide angle visual display. Their results revealed that performance was effected the most in the longest lag condition, but there was no statistical difference in sickness incidence rates between the delay conditions. However, the investigators advised caution in generalizing the results because only two types of tasks were examined and there was no inertial motion platform (Uliano, Lambert et al., 1986). In particular, they suggested that the results could be different for other types of tasks or, in a motion-base simulator, if a lag between the visual and inertial systems was present.

Frank, Casali, and Wierwille (1988) evaluated visual-motion coupling delays and cuing order in a driving simulator using different combinations of transport delays (0, 170, or 340 ms) in either the visual system, motion system, or both systems. Their results showed that zero delay in either system was the most desirable condition, whereas delays in the visual or motion system

increased participant's overall severity of simulator sickness. However, visual delays effected sickness incidence more than motion system delays. When asynchronous delays occurred between the visual and motion systems, sickness was greater when the motion system led the visual system. In contrast, Padmos and Milders (1992) cited research findings which indicated that the *visual* imaging system should not have a time lag with respect to the inertial system.

The general recommendation for reducing the potential for sickness due to cue asynchrony is to limit the delay between any two system cues to no more than 35 ms (Lilienthal as cited in Pausch, Crea, & Conway, 1992). Kennedy, Berbaum et al. (1987) also recommended that lag in the motion base should not exceed 83-125 ms and there should be no more than 40 ms asynchrony between visual and inertial cues.

SSQ Profile Analysis Studies

Obviously, the equipment of a VE system is what creates the simulated environment and previous research has identified equipment features as one of the major factors influencing VE sickness. Nonetheless, there is still limited knowledge concerning the effects of VE system design variables on sickness in general. Moreover, there is a paucity of literature that addresses the relationship between equipment features of VE systems and the specific types of sickness symptoms that are induced by exposure to them.

As previously mentioned, the most frequently used measurement technique to assess the signs and symptoms of sickness in various provocative environments, particularly simulator and VR systems, is the Simulator Sickness Questionnaire (Kennedy, Lane et al., 1993). Kennedy and Fowlkes (1992) suggested that like motion sickness, the polysymptomatic nature of sickness induced by exposure to VE systems was advantageous because differences in symptoms could

provide diagnostic information regarding the source of the symptoms. Accordingly, Lane and Kennedy (1988) suggested that the subscale measures of the SSQ could be used to provide more precise information about the particular systems of the body which were affected by a provocative motion stimulus. In particular, the authors recommended that the SSQ subscale scores should be used in studies designed to compare the effects of different motion environments or studies investigating the causality of sickness attributable to different aspects of the stimulus.

Relatedly, Kennedy, Drexler, Stanney, and Harm (1997) indicated that similarities in SSQ symptom profiles from two different environments would suggest a common cause, even if the similar profile occurred in a different VE with different visual display systems or other design characteristics. Likewise, the authors suggested that SSQ profile differences (e.g., excessive visual disturbance) may signal differences in specific equipment design features that differentially affect the severity and types of symptoms reported. Theoretical support for these hypotheses is provided by the psychophysical linking hypothesis, the concepts of endophenotypes and surrogate measures. An overview of each of these concepts is provided in the following sections.

Psychophysical Linking Hypothesis

In vision research, psychophysical experiments are typically conducted in order to relate the results to the underlying physiological processes of the visual system (Boynton & Onley, 1962). Accordingly, Brindley (1960) proposed a psychophysical linking hypothesis that could be used to relate physiology and psychophysics. His theory suggested that if a physiological hypothesis about a particular function is postulated to explain a given result from a sensory

experiment, then the theory must also include hypotheses containing psychological terms (i.e., a psychophysical linking hypothesis).

First, Brindley (1960) distinguished between two types of observer judgments involved in psychophysical experiments: Class A observations (those that produced the same sensation) and Class B observations (those that involved more complex experiences). Specifically, Brindley defined Class A observations as those where two physically different stimuli under a particular set of conditions produced the same sensory experience (i.e., the same psychophysical judgment) and those where the stimuli (i.e., the two physically different stimuli), under a different set of conditions produced a different sensory experience. In contrast, Class B observations require the observer to abstract the quality of the psychological visual characteristic (e.g., brightness, hue) of interest from a complex visual experience (Brindely, 1960). Consequently, Brindley considered Class A observations as superior to those of Class B and thus recommended that Class B observations should be converted into Class A observations when possible in order to relate the data from psychophysical experiments to physiological hypotheses (Boynton & Onley, 1962).

Brindley (1960) then proposed a psychophysical linking hypothesis for Class A observations which stated that "whenever two stimuli cause physically indistinguishable signals to be sent from the sense organ to the brain, the sensations produced by these stimuli….must also be indistinguishable" (p. 146). In other words, physically different stimuli may produce the same signal that creates an identical sensory experience (Boynton & Onley, 1962). Brindley also noted that while the hypothesis was the most general theory that had been proposed, it was also "the most difficult to doubt" (p. 146). Boynton and Onley (1962), however, criticized Brindley's implied application of the psychophysical linking hypothesis which suggested that experiments involving Class A type observations would be capable of testing a psychophysical linking

hypothesis. In contrast, the authors suggested that an experiment involving Class A observations may generate physiological data that could be used to relate psychophysics and physiological theory through a Class A converse of the psychophysical linking hypothesis (CPLH). Boynton and Onley's CPLH stated that "whenever the sensations produced by two [stimuli] are subjectively indistinguishable…one may conclude that the stimuli which produced these sensations caused physically indistinguishable signals to be sent from the sense organs to the brain" (p. 385). That is, the same sensory experience may be caused by two stimuli that produce an identical sensory signal.

Additionally, Boynton and Onley (1962) indicated that testing a specific experimental procedure for adherence to a Class A criterion (i.e., two stimuli produce an identical sensory experience) actually involves trying to prove the null hypothesis. Traditional statistical tests of hypotheses are analogous to proof by contradiction where the theory the researcher wants to prove, or support, is defined as the alternative hypothesis and the contradictory theory is the null hypothesis (Mendenhall & Sincich, 1995). Then, the result of the statistical test would indicate whether to reject or fail to reject the null hypothesis with a known probability of a Type I error (i.e., rejecting the null when it is true) denoted by α. Moreover, in this type of hypothesis test a failure to reject the null hypothesis would not imply that the null was true because the probability of a Type II error (i.e., not rejecting the null when it should have been rejected), designated as β, was unknown (Mendenhall & Sincich, 1995). In contrast to this type of traditional hypothesis test, when the null hypothesis is selected as the theory the researcher wants to support (i.e., testing the null hypothesis) and the results indicate that the null should not be rejected (i.e., the data support the theory), different values of ß would then have to be investigated for specific

alternatives, which Mendenhall and Sincich (1995) noted is a "tedious and sometimes extremely difficult task" (p. 436) that should be avoided if possible.

Finally, Boynton and Onley (1962) questioned the validity of Brindley's dichotomy of Class A versus Class B observations by identifying several instances where Class A observations were involved in a psychophysical experiment, but where the Class A psychophysical linking hypothesis was not considered acceptable. In general, the authors suggested that 'definitely Class A' observations could be distinguished from "definitely not Class A' observations, but the classification of observations from psychophysical experiments existed along a continuum. Therefore, the authors proposed that Brindley's dichotomy should be expanded to six classes of observations: three types of Class A observations and three types of Class B observations. Boynton and Onley noted, however, that relating data from psychophysical experiments to physiological theory will always be somewhat tenuous due to the nature of psychophysical experimentation. Specifically, the authors explained that due to the uncertain relation between physiological events and conscious experience, most psychophysical experiments attempt to test the truth of a psychophysical linking hypothesis while simultaneously assuming that the hypothesis is true in order to examine the quantitative nature of the hypothesis through the use of psychophysics.

Surrogate Measures

In the area of performance measurement, Lane, Kennedy, and Jones (1986) noted that operational (i.e., real-world) performance measures generally suffered from low reliability and thus, such measures were insensitive to performance changes. To overcome the lack of reliability in field measures, the authors proposed that a set of highly reliable measures, which

are separate from the real-world operational criteria (i.e., outside of the direct task context) but similar in terms of the particular skills required, could be developed and used as an alternative to the operational measure (c.f. also Kennedy, Lane, & Kuntz, 1987). The Lane et al. (1986) surrogate concept suggested that the alternative tests, called surrogate measures, "are related to or predictive of a construct of interest (such as "true" field performance), but are not direct measures of that construct" (p. 1400). Accordingly, performance on combinations of simpler, typically uni-dimensional tests that are designed to tap the elementary components underlying more complex performance, could be used to predict large portions of the variance on complex, multi-functional tasks (Turnage & Lane, 1987). Moreover, because the reliability of a surrogate measure is much greater, it would be logically and statistically reasonable to expect that it may predict more of the true variance in the criterion performance of interest (Kennedy, Lane, & Kuntz, 1987; Turnage & Lane, 1987). In addition to high reliability, two other important characteristics of surrogate measures are they should correlate with the real-world performance construct and be sensitive to the same factors that would affect the overall performance criterion (Kennedy, Lane, & Kuntz, 1987; Kennedy, Turnage, & Lane, 1995, 1997; Lane et al., 1986). Surrogate measures only evaluate performance on components (or factors) which are common to the performance measure; they do not need to involve specific operations in common with the performance criterion (Lane et al., 1986). Thus, surrogate measures are separate from the task performance itself.

Computerized surrogate tests have been shown to be stable and reliable performance indicators in a study on the prediction of complex flight performance in a flight simulator (Turnage, Kennedy, Gilson, Bliss, & Nolan, 1988) and various studies on the effects of different stressors (blood alcohol levels, chemotherapy, hypoxia, sleep loss, etc.) on performance

(Kennedy, Lane, & Kuntz, 1987; Kennedy, Turnage, & Lane, 1995, 1997). Kennedy, Turnage, and Lane (1995, 1997) described a series of studies that used surrogate measures to link laboratory performance scores to real-world performance. In the first study, a dose-equivalency relationship was determined where performance deficits on surrogate tests which sampled all factors of basic information processing were related to graded dosages of alcohol (i.e., blood alcohol level [BAL]). The purpose of the study was to develop a composite score for the surrogate tests and to establish prediction equations for BAL in which performance decrements on the surrogate tests could be used to predict BAL. Because random variability in performance scores across trials on individual tests can weaken or mask trends in performance changes, the individual surrogate tests were combined into a single composite score to reduce random variability and thus stabilize the relevant variance (Kennedy, Turnage, & Lane, 1997).

The second study evaluated standardized intelligence and aptitude tests which were known to be valid predictors of real-world (operational) performance (Kennedy, Turnage, & Lane, 1995, 1997). The standardized tests included two IQ tests: the Weschler Adult Intelligence Scale Revised (verbal, performance, and full-scale IQ) and the Wonderlic Personnel Test (quick estimate of full-scale IQ) and two aptitude tests: the American College Testing (ACT; measure of performance potential in college) and the Armed Services Vocational Aptitude Battery (ASVAB; extensive aptitude test). Surrogate measures of aptitude and intelligence were developed by determining the predictive relationship between the surrogate tests and the standardized tests (Kennedy, Turnage, & Lane, 1995, 1997). Each surrogate test was correlated with each standardized test and a composite equation was then developed for each standardized test. Specifically, regression equations of surrogate scores were developed that predicted scores on the standardized tests. Finally, the surrogate equations were applied and

cross-validated in two separate alcohol studies in order to determine the relationship (i.e., correlation) between the actual and predicted performance decrement scores due to alcohol consumption (Kennedy, Turnage, & Lane, 1995, 1997). Accordingly, performance decrements on the surrogate tests associated with a particular blood alcohol level could be related to performance decrements on the standardized tests.

Lane, Kennedy, and Jones (1986) explained that it is possible for performance changes on surrogate tests, as with any performance test, to be mediated by variance that is not related to the criterion. Consequently, if degraded performance on a surrogate test occurs from exposure to a particular stressor, it cannot be definitively *proven* that the criterion performance would also have been degraded by the same stressor. However, because surrogate measures, when selected, are already shown to be sensitive to the same factors as the criterion and are correlated with the performance construct of interest, the authors indicated that performance changes on surrogate tests "which tap skill components in common with the operational tasks would constitute presumptive evidence for similar changes on the criterion" (p. 1401).

Endophenotypes

In the area of psychiatric genetics, an endophenotype-based approach has been suggested as a potential aid in diagnosis of psychiatric disorders (Gottesman & Gould, 2003). Psychiatric diseases are currently classified on the basis of the overt phenotypes (i.e., the observable behaviors of an individual) that are characteristic of the particular disease. However, the genetics underlying psychiatric diseases such as schizophrenia are inherently complex due to the complexity of the human brain and the multifactorial and polygenic origins of the diseases (Gottesman & Gould, 2003). The authors pointed out that genetic dissection of the diseases in

order to definitively identify specific genes or gene regions involved in the development of the diseases has not been very successful because the diagnostic classification scheme is not based on measures of the underlying genetics of the disease. Consequently, Gottesman and Gould discussed the concept of endophenotypes, which involve "measurable components unseen by the unaided eye along the pathway between disease and distal genotype", as a method that could be used to fill the current gap between the genes and the disease processes (Gottesman & Gould, 2003, p. 636). Phenotypes are the observable characteristics of an organism that are the result of both genetic and environmental influences whereas endophenotypes are the internal phenotypes discoverable by microscopic examination (Gottesman & Gould, 2003). Other terms that have been used to refer to endophenotypes include intermediate phenotype, subclinical trait, vulnerability marker, and biological marker.

Gottesman and Gould (2003) indicated that the pathway of interest in the diagnosis of psychiatric disorders leads from the genes (genotypes) to the behavioral macros (phenotypes), and the endophenotypes (intermediate variables) provide a link between the two (i.e., they mark the path between genes and the behavior of interest). They suggested that diagnosis of psychiatric disorders, which have complex genetic underpinnings, could therefore be improved by tapping into the endophenotypes in the pathway to the disease. Specifically, an endophenotype-based approach would facilitate diagnosis of psychiatric disorders by decomposing the disease syndrome into simpler components (endophenotypes) that could provide more straightforward analysis of the genetic basis of the disorder of interest (Gottesman & Gould, 2003). Moreover, endophenotypes would assist in the identification of aberrant genes that make an individual vulnerable to a particular psychiatric disorder by providing a means to

identify the "downstream traits or facets of clinical phenotypes as well as the upstream consequences of genes" (p. 637).

The criteria for a candidate endophenotype include an association with the disease in the population, inheritability, and presence in an individual even when the disease is not active (Gottesman & Gould, 2003). The methods used to analyze endophenotypes include measures of cognitive, neurophysical, neurophysiological, neuropsychological, biochemical, and endocrinological functioning using tools such as self-report, functional magnetic resonance imaging (fMRI), and positron emission tomography (PET). These types of measurements not only represent objective, quantifiable methods for disease diagnosis, they also constitute candidate endophenotypes that may represent the primary inclusion/exclusion criteria for genetic linkage studies.

As an example of the endophenotype concept, Gottesman and Gould (2003) described research on candidate endophenotypes for schizophrenia, a complex psychiatric disorder that involves a wide spectrum of behaviors and experiences. The authors noted that the source of the disorder is the individual's genes, but other influences such as the environment also play a role in determining the behavioral macros (i.e., phenotypes) typical of the disorder. Several different genes and gene regions have been identified which are known, or suspected, to be involved in schizophrenia (Gottesman & Gould, 2003). These genes and gene regions have been linked to a variety of more specific functions (i.e., candidate endophenotypes) such as working memory, oculomotor function, glial cell abnormalities, and sensory motor gating. For example, neuropsychological research has revealed deficits in sensory motor gating in schizophrenia patients. Specifically, neuropsychological tests (e.g., prepulse inhibition of the startle response and P50 suppression) have shown defects in inhibitory neuronal circuits in schizophrenic patients

compared to healthy subjects (Gottesman & Gould, 2003). Other research studies using these tests have also identified gene and chromosomal regions which may be involved in this particular candidate endophenotype. On the behavioral level, the presence of this candidate endophenotype has been exhibited in schizophrenic patients' reports of difficulty filtering information from multiple sources. Thus, the sensory motor gating endophenotype provides a link between the genetic basis of schizophrenia and the "upstream consequences of the genes" (i.e., one of the behaviors of the schizophrenic patient).

Although the psychophysical linking hypothesis, endophenotypes, and surrogate measures are concepts from diverse areas of the literature, they all have similar underpinnings. First, all of the concepts are based on the idea that direct measurement of a criterion of interest is not always feasible and instead, information on measurements obtained from one construct can be used to make inferences about another construct. Specifically, they use the results of a particular measure as a bridge or link between a simpler or more elementary constructs and a more complex construct. These three concepts also indicate that if changes occur on the measured construct, it is reasonable to presume that similar changes occur on the criterion of interest. Moreover, the psychophysical linking hypothesis argues that conclusions about different stimulus conditions can be drawn from empirical findings of the response (i.e., similarities or differences in the response correspond to similarities or differences in the stimuli). Accordingly, support for the hypotheses postulated by Kennedy, Drexler, Stanney, and Harm (1997) regarding similarities and differences in symptom profiles as being suggestive of common or different causes are theoretically supported by the literature on the psychophysical linking hypothesis, endophenotypes, and surrogate measures. Specifically, in the current research the SSQ will be used to measure sickness (i.e., a simpler measure of a complex construct) and

differences in the SSQ subscales will be used to relate differences between VE systems to the equipment features of the systems (i.e., hypotheses about the equipment will be used to explain differences in sickness scores).

The following sections provide a summary of different investigations that have used the SSQ subscales to evaluate the effect on sickness symptoms related to some aspect of the VE equipment. There are, of course, other studies available in the open literature where data from the SSQ subscales were reported. However, these other studies employed different independent variables (e.g., kinematics, scene complexity, scene detail, gender, motion sickness history, and motion sickness medication) rather than the characteristics of the equipment.

System Dependent Studies

Various laboratory experiments have used the SSQ to investigate the effects of specific VE system design features on sickness. These studies have typically involved systematic manipulations of various equipment features (e.g., transport delay, field-of-view) in order to determine the relative contribution of the specific feature to the incidence of sickness. However, in the majority of the studies reviewed, the authors only reported the effects on overall sickness (i.e., the SSQ Total Severity [TS] score). For example, Lin et al. (2002) used the TS score to report the effects of different field-of-view sizes on sickness, Draper et al. (2001) evaluated the effects on the TS score for two system time delays, and Lampton, Rodriguez, and Cotton (2000) investigated changes on TS score during simultaneous exposure of two-person teams using multiple position tracking sensors. Accordingly, the studies cited below specifically reported the effect of different aspects of VE equipment on the three SSQ subscale scores including the type

of visual display, position tracker delay, sensory feedback devices, simulator platform, motion frequency, and various HMD-based VR system configurations.

Display Type

Using a repeated-measures design to investigate performance on a target location task, Garris-Reif and Franz (1995) also evaluated the effects on sickness induced by a conventional desk-top display (i.e., 21-inch color monitor) and a head-slaved VR display (i.e., a BOOM). There was a significant difference in sickness severity between the two system configurations where the VR display produced more severe sickness symptoms. The SSQ subscales also showed similar results; scores on the subscales for the VR display were greater than scores for the desk-top display. An analysis of the subscale scores for the VR display indicated that Oculomotor disturbances were the most pronounced effect, followed by the Disorientation and Nausea subscales. Garris-Reif and Franz suggested one potential explanation for the difference in sickness incidence was the difference in the FOV between the two displays. In particular, the FOV for the VR display was approximately 140° whereas the desk-top FOV was only about 40°.

Häkkinen, Vuori, and Puhakka (2002) compared the sickness symptoms produced by watching a 2D movie with an HMD and playing a racing game with the same HMD in stereoscopic mode. The results of their study showed significance differences between the two conditions on all three of the SSQ subscales. However, since a different task was used for the two conditions, the results cannot be interpreted in terms of symptom differences induced by viewing 2D and 3D images in an HMD display.

Position Tracker Delay

A study by Nelson, Bolia, Roe, and Morely (2000) investigated the effects on sickness as a function of position tracker delay (46, 96, 146 msec) in a see-through HMD display system. The results of their study indicated that SSQ scores for Total Severity and the Nausea and Oculomotor subscales varied with exposure duration, but none of the SSQ scores were affected by the time delay.

Sensory Feedback

Jaeger and Mourant (2001) investigated differences in sickness symptoms associated with the mode of locomotion (either static or dynamic) in an HMD-based VR system. In the dynamic condition, participants walked on a manually-powered treadmill and thus, physical activity was required to move through the VE. In contrast, participants in the static condition controlled their movement through the VE using a hand-controlled device (i.e., computer mouse). The authors hypothesized that the dynamic locomotion condition would produce less sickness symptoms because whole-body movement would stimulate the vestibular system which would reduce the potential for conflict between the visual and vestibular systems. As expected, an analysis of the SSQ data showed that sickness severity was significantly greater in the static locomotion condition. Moreover, their analysis revealed differences in the SSQ profiles for the two conditions. The profile for the static locomotion condition showed that scores on the Disorientation subscale were greater than the Oculomotor subscale which was greater than the Nausea subscale (i.e., a D>O>N profile). However, the dynamic condition profile revealed that Oculomotor symptoms were greater than Nausea, followed by Disorientation (i.e., a O>N>D

profile). Therefore, Jaeger and Mourant concluded that the vestibular feedback associated with the physical activity in the dynamic condition reduced the severity of sickness.

Fowlkes et al. (2002) cited evidence from other empirical research that indicated posture during and after VE exposure could be stabilized through light touch to the index finger. Therefore, the authors used the SSQ to investigate whether haptic input, using a haptic glove to provide tactile stimulation (vibration) to the index finger, could affect sickness induced by exposure to VR systems. Accordingly, their empirical study compared two types of feedback, auditory and tactile, during a target acquisition task. A series of virtual targets were presented via an HMD and the participants task was to "touch" the target as quickly and accurately as possible. When contact was made with the target, participants in the auditory condition heard an "impact" tone whereas the haptic participant group received feedback through light vibration to the index finger. The results of the study revealed that scores on the Oculomotor subscale were the most prominent in both condition, but a significantly lower incidence of Nausea-type symptoms was found in the haptic condition.

Simulator Platform

Curry, Artz, Cathey, Grant, and Greenberg (2002) evaluated the effects of exposure to a fixed-base and motion-base driving simulator on SSQ scores. The fixed-base system was a high fidelity simulator that used three projectors to display the driving simulation and provided a 140° x 25° FOV. In contrast, the motion-base system used five projectors to display images on the inside surface of a 24-ft. dome mounted on a six degree-of-freedom (DOF) platform. The FOV for the motion-base simulator was 300° (180 x 40 in the front, 120 x 27 in the rear). A similar driving scenario (2-lane expressway) and exposure duration (~ 50 min.) was used in both

systems. An analysis of the SSQ data showed a significant difference in the total severity of sickness. Although the severity of sickness was significantly greater in the fixed-base simulator, the two systems appeared to have relatively consistent SSQ profiles (i.e., D>O>N).

Motion Frequency

Kennedy, Drexler, and Compton (1997) examined the motion characteristics in two motion-base flight simulators where sickness had been reported in order to determine whether the sickness was related to the motion of the simulator. Motion data from accelerometers placed in the simulators during a one-hour simulation were compared for the 2F64C, a rotary wing (i.e., helicopter) simulator with considerable reports of sickness and the 2F87F, a fixed wing (aircraft) simulator with minimal reports of sickness. The authors hypothesized that motion-based simulators which reported significant levels for the SSQ Nausea subscale would also have significant amounts of energy in the 0.2 Hz region. The results of the motion spectra analyses indicated that, as expected, the 2F64C helicopter simulator with considerable reports of sickness generated high levels of inertial motion in the 0.2 Hz region (high levels of acceleration in g_z and g_y) whereas the 2F87F fixed wing simulator with minimal sickness incidence had a negligible amount of acceleration in the 0.2 Hz range. The authors concluded that the very low frequency motion in the helicopter simulator contributed to some, but not all, of the reported simulator sickness.

VR System Configuration

Cobb, Nichols, Ramsey, and Wilson (1999) reported the results from a series of experiments with different VR system configurations (HMDs, computer processor speeds,

tracker delays, etc.) where differences in sickness symptoms were found between experimental conditions. In particular, their results revealed similar SSQ profiles for studies where the same VR system configuration was used. In contrast, a comparison of SSQ scores between a high-end HMD display and a low-end HMD (i.e., smaller visual display with low resolution) showed the low-end system produced higher scores on the Oculomotor subscale. In another study, the speed of the computer processor varied but the same HMD was used in both conditions. The slower processor speed was associated with greater lag in the display update rate. In this experiment, the results indicated that the slower processor speed resulted in higher scores on the Disorientation subscale. Finally, a comparison of the sickness scores based on the type of display revealed that use of an HMD display provoked higher scores on all three of the SSQ subscales compared to viewing the same stimulus on a CRT monitor.

While the system-dependent experiments cited above can be useful for answering questions about specific design features for a given VE system, the results are generally limited to the particular VE device under investigation. In contrast, preliminary non-system specific research has been conducted to evaluate similarities and differences in SSQ symptom profiles related to system design features.

System Independent Studies

Although the relationship between equipment design features and the particular symptoms elicited by those features have yet to be completely identified, Lawson et al. (2002) noted that "important steps in this direction have been taken by Kennedy and colleagues" (p. 599). Kennedy, Jones, Lilienthal, and Harm (1994) first suggested that the three SSQ subscale scores could be used as a system profile. Specifically, the authors asserted that comparisons of

the SSQ profiles could provide information about the nature of the sickness reported in a particular device as well as diagnostic information about the characteristics of the equipment that may differentially affect reported sickness. They argued that similar symptom profiles between devices could imply similar causes of sickness whereas profile differences may signal different causes. To support their hypothesis, the authors compared the SSQ profiles obtained from different military flight simulators as well as from a few other provocative environments. First, a comparison of the SSQ profiles from four helicopter simulators was presented; one pair of identical simulators located in the same city and another pair of identical simulators located in different cities. *Within* each pair of profiles, the two identical simulators appeared to have mirror images of each other, whereas the profiles differed slightly *between* the two different pairs of simulators.

Kennedy, Jones et al. (1994) also presented separate SSQ profiles from several Army and Navy helicopter simulators. Again, a comparison of the profiles suggested that while the level of severity for the three SSQ subscales differed among the simulators, the overall *profile* for the helicopter simulators were similar. The profile for the helicopter simulators indicated that the Oculomotor subscale was the largest symptom factor, followed by the Nausea subscale, and then Disorientation (i.e., O>N> D). The SSQ profiles from other provocative environments where motion sickness-like symptoms have been reported were then presented and compared to the simulator profiles. The profile from the Pre-flight Adaptation (PAT), a VR system that is used by NASA to train astronauts in the illusory experiences that will occur while in space, and the profile of space motion sickness symptoms reported by NASA astronauts were compared to the profile of the average sickness from a dozen Navy flight simulators. The authors reported that the two NASA environments (i.e., PAT and space sickness) produced profiles similar to each

other; Nausea was greater than Disorientation, which was greater than Oculomotor discomfort (N>D>O). However, the two space sickness profiles were different from the profile of simulator sickness in Navy flight trainers (i.e., O>N>D). Kennedy, Jones et al. also compared the SSQ profiles from three different VR systems to the NASA-PAT system. This comparison indicated that while the magnitude of the three SSQ subscale scores differed, the profile for two of the VR systems resembled the NASA-PAT profile (i.e., N>D>O), but the third VR system exhibited a slightly different profile (O>D>N). Consequently, the authors suggested that a comparison of the equipment features of the systems could be used to reveal the underlying cause of the similarities and differences in the SSQ profiles.

Kennedy, Drexler, Stanney, and Harm (1997) presented and compared the SSQ profiles from several different forms of motion sickness. The SSQ profile for seasickness indicated that the Nausea component was greater than the Oculomotor factor, which was greater than Disorientation (i.e., N>O>D profile). In contrast, the profile of space sickness was characterized by a significant amount of Nausea and Disorientation, but relatively little Oculomotor disturbances (i.e., N>D>O). The profile of simulator sickness showed a significant amount of Oculomotor disturbance, but less Nausea and Disorientation (i.e., O>N>D). Kennedy, Drexler et al. (1997) also offered support for their hypothesis that VE devices can have a specific SSQ profile or "signature" by presenting a comparison of sickness profiles from different simulator and VR systems. First, profile comparisons were reported for five different military flight simulators. All of the simulators were motion-based Navy and Marine Corps rotary wing (helicopter) simulators that employed multiple CRT displays to present the computer generated imagery. Two of the simulators, manufactured to the same specification, but located at geographically different training installations, appeared to exhibit a very similar profile;

Oculomotor disturbance was the highest subscale score with relatively lower Nausea and Disorientation scores. Similarly, two other simulators, each manufactured to the same specification, also appeared to exhibit a similar profile to each other, but slightly different than the other two devices that were manufactured to a different specification. In particular, the SSQ profile for the two simulators revealed a higher Nausea component, but lower Oculomotor and Disorientation scores. Additionally, Kennedy, Drexler et al. (1997) indicated that nine out of ten Navy and Army helicopter simulators, all of which had a motion-base and multiple CRT displays but different simulated missions or tasks, exhibited a similar pattern of SSQ profiles scores. The authors also pointed out that the overall sickness incidence (i.e., Total Severity score) was different among all of the simulators.

Next, Kennedy, Drexler et al. (1997) provided the sickness profiles from experiments with four different HMD VR systems. Although the systems differed in terms of the dynamics and displays of the HMDs as well as differences in scene content, the authors indicated that the SSQ profiles were relatively consistent across the different systems. The profiles, in general, exhibited higher Disorientation-type symptoms than Nausea symptoms and Oculomotor symptoms were the lowest (i.e., D>N>O). Finally, the SSQ profile for flight simulators was compared to the profile for VR systems. The authors noted that the majority of the simulators showed Oculomotor disturbances as the largest factor and Disorientation as the weakest sickness contributor whereas the VR system profile showed the reverse configuration (i.e., Disorientation was the highest category and Oculomotor disruption was the lowest). The authors, therefore, concluded that the differences in SSQ profiles implied that there were differences between the sickness induced by exposure to VR and simulator systems. In other words, the two types of VE systems may produce different forms or types of sickness (cf. Table 1 for a summary of the SSQ

profiles). However, they also indicated that whether the SSQ profiles generalize beyond a few devices and whether the profiles can be used as an aid to determine the causes of sickness in different systems still must be determined.

Table 1. SSQ Profiles from Various Provocative Motion Environments

Environment	Profile
Sea	N O D
Space	N D O
Flight Simulator	O N D
HMD VR System	D N O

Stanney, Kennedy, and Drexler (1997) also emphasized the importance of determining whether the sickness induced by exposure to VR systems is similar to the sickness induced in simulator systems. The authors suggested that differences in SSQ sickness severity and/or symptomatology could indicate that simulator sickness and VR sickness (cybersickness) are distinct types of motion sickness. In a comparison of the SSQ Total Severity (TS) scores across eight different VR experiments and ten military flight simulators, the authors reported that the average TS score for VR users was approximately three times greater than the average severity reported by flight simulator users. Differences in the SSQ profiles for VR systems and military flight simulators were again highlighted.

Some researchers could argue that the differences in sickness between the two types of systems were due to differences between the user populations (Stanney et al., 1997). The authors explained that the simulator users were mainly male military aviators that were essentially self-selected as resistant to motion sickness whereas the VR users included approximately equal numbers of male and female college students that were not pre-selected for their resistance to motion sickness. Although Stanney et al. (1997) acknowledged that the population differences

could have been a contributing factor to the differences in reported sickness, they suggested that the primary factor in the SSQ profile differences was the result of fundamental differences in the stimulus provided by VR and simulator systems. In other words, because the pattern of SSQ symptoms produced by VR systems was different than simulator systems, the sickness experienced by VR users may be caused by different factors than the sickness experienced by flight simulator users. Therefore, the authors concluded that cybersickness and simulator sickness appear to be distinct forms of motion sickness.

Overall, the previous literature provided information about similarities and differences in SSQ profiles which suggests the profiles may contain important diagnostic information about the cause of sickness experienced in different systems. However, this research must be considered speculative because analyses of the SSQ data only involved visual comparisons of the symptom profiles. Conversely, there has only been one report to date where non-system specific quantitative analyses of the SSQ subscales were attempted, which used discriminant and chi-square analyses (Kennedy, Drexler et al., 2003). The goal of the discriminant analyses was to determine how well the SSQ subscales predicted group membership on various equipment characteristics. Therefore, the authors separated the SSQ data into different groups, or "classes", based on binary features of the equipment (e.g., motion-base vs. fixed-base simulators, binocular vs. bi-ocular HMDs, simulators vs. VR devices, etc.). Next, one set of scores was created for each device analyzed which consisted of the average scores for all participants in each study. Although the results for the individual comparisons were not reported, Kennedy, Drexler et al. indicated that the result for the simulator versus VR device comparison revealed 'strong' results. The authors noted however, that it was unknown whether the result of the simulator-VR

comparison was due to differences in the SSQ subscales or differences in symptom severity (i.e., magnitude) between the two types of systems.

Kennedy, Drexler et al. (2002) also conducted a chi-square test on the SSQ data for each study in their large database. The purpose of the chi-square analyses was to determine how well the SSQ profile for the overall study matched the profile for each participant in the study. Although there were six possible permutations of the three subscales, the authors used only three of the possible SSQ profiles (based on which SSQ subscale [N, O, or D] was the highest) for each analysis. Additionally, the SSQ data were trimmed to exclude those participants that reported no symptoms (where the expected value would be 1/6 or 16½%). The results of the chi-square tests showed that a participant's profile matched the profile for the overall study, in terms of the highest SSQ subscale, for 60% of the simulator studies and 50% of the VR studies. It should be noted that the data from the studies where the individual profiles did not match the overall study profile indicated that 83% of the simulator studies and 22% of the VR studies contained relatively small sample sizes (≤ 25 participants). Consequently, additional quantitative research is still needed in order to identify the relationship between equipment design features and the particular types of symptoms elicited by those features.

Significance of the Research

In order for the science and technology of VE systems to be practical, the various causes of the physiological effects associated with use of the systems must be fully understood. In the past, the research community focused a great deal of attention on the identification and examination of several factors that influence the incidence and severity of sickness, particularly individual user characteristics, exposure duration, and usage schedule. Obviously, the equipment

creates the artificial environment and previous research has identified equipment features as one of the major factors influencing VE sickness. Nonetheless, there is a paucity of literature that addresses the relationship between equipment features of VE systems and the resultant side effects. Moreover, of the five major determiners of sickness discussed previously, manipulation of VE equipment features can provide the most direct, practical, and economical means to controlling sickness. Specifically, Kennedy (1996) pointed out that the characteristics of individual users can only provide indirect control of sickness through careful selection of the individuals exposed to a VE system, which will prohibit a significant proportion of potential users from reaping the benefits of the technology. For entertainment and/or some research applications, exposure duration can be directly manipulated in order to minimize sickness, but is likely to hinder effective use of VE technology as a training tool, particularly for applications which require prolonged immersion in the simulated environment. Similarly, direct manipulation of the usage schedule can be used to facilitate adaptation to sickness, although this approach can be expensive (e.g., labor costs for the user, trainer, and equipment operators, decreased operational readiness until adaptation is achieved, etc.) and is not always effective for controlling sickness (i.e., some users may never adapt). Finally, neither direct nor indirect control over kinematics can be achieved due to the interactive nature of VE systems and therefore, can only be measured for use as covariates in empirical research. However, an understanding of the physiological effects of equipment features on users can be used to identify the specific features that should be targeted for redesign which could provide the most effective approach to solving (or at least minimizing) the sickness associated with exposure to VE systems. Thus, it is essential that human factors engineering research be devoted to

understanding the differential effects of various equipment design features on sickness outcomes in order to facilitate effective management of VE-induced sickness (i.e., minimize side effects).

Different types of sickness symptoms (e.g., nausea, oculomotor disturbance) will generally require distinctive technological solutions to reduce the occurrence of these symptoms. Until the equipment features that influence specific types of symptoms are identified, technological solutions to the sickness problem, such as engineering modifications of equipment features that contribute to sickness, cannot be achieved. Stanney, Mourant, and Kennedy (1998) pointed out that "it is essential that VE developers ensure that advances in VE technology do not come at the expense of human well-being" (p. 339). However, without an understanding of the relationship between system design features and sickness outcomes, some technological advances in VE systems will be inconsequential or worse, may amplify the sickness problem. Relatedly, the research on VE sickness conducted to date has provided some general recommendations for reducing the side effects of VE exposure (e.g., Stanney, Kennedy, & Kingdon, 2002) but, there are currently no specific guidelines for equipment design to minimize sickness. The results of the research will address this deficit by identifying equipment features of VE systems that significantly influence the SSQ symptom subscales (profiles) which could then be used to specify potential technological solutions to minimize sickness.

Several laboratory experiments have been conducted to investigate the effects of specific system design features on sickness. These empirical studies of different system engineering features typically have involved systematic manipulations of various equipment features (e.g., transport delay, field of view, computer processing speed) to determine the relative contribution of the system feature to the incidence of sickness. While these types of system-dependent experiments are useful for answering questions about specific design features for a given VE

system, the results are generally limited to the particular VE device under investigation. Additionally, most of the studies only evaluated the effects of the manipulated variable on overall sickness. There is limited research on the system design features that influence specific types of symptoms. In contrast, preliminary non-system specific research has been conducted to evaluate similarities and differences in SSQ symptom profiles related to system design features. However, this research is speculative because analyses of the sickness data only involved visual inspection of the symptom profiles. To date, non-system specific quantitative research relating sickness symptoms (profiles) to VE system design features has not been conducted. Thus, the design features that are best suited to minimize particular types of symptoms related to VE exposure are still an open question. Consequently, this research will afford a deeper understanding of the relationship between the engineering characteristics of different VE systems and specific types of sickness symptoms so that specific design recommendations for equipment design can be developed.

Kennedy and Fowlkes (1992) argued that large numbers of subjects are crucial to reveal a significant treatment effect on sickness, particularly an effect of different equipment features. Similarly, Kennedy, Drexler, Stanney, and Harm (1997) suggested that if sufficient SSQ data were available, it may be possible to identify a consistent symptom configuration of the three SSQ subscales (i.e., an SSQ profile) within a given VE device as well as differences in SSQ profiles between VE devices. Accordingly, the lack of non-system specific research related to the effects of system design features on different types of symptoms is most likely due to limited access to sickness data from a large group of different VE devices. A substantial amount of SSQ data collected from a diverse set of VE systems was available for this research. The research therefore offers a unique opportunity to evaluate the incidence and severity of VE sickness

across a broad spectrum of simulators (flight and driving) and VR devices. Specifically, access to the SSQ data collected from a wide range of VE systems will permit an evaluation of the similarities and differences in profiles WITHIN a given type of VE device (i.e., different HMDs and simulators) as well as analyses of profile differences BETWEEN device types. Analyses of the large SSQ database will also provide quantitative evidence that will either support or refute the assumption that simulator sickness and cybersickness are distinct forms of motion sickness.

Research Implications

The proposed research will significantly contribute to the development of human factors guidelines for the design of VE equipment by providing general (i.e., non-system specific) VE design principles that will reduce the side effects associated with exposure to different VE devices. Logically, different types of side effects (e.g., disorientation, nausea) can require distinctive technological solutions to reduce their occurrence. Although simulators and VR devices both provide visually interactive computer-generated environments, two different terms, simulator sickness and VR sickness (or cybersickness) have been used to distinguish between the adverse effects produced by the two types of VE systems. If statistically significant differences are found between the SSQ profiles for simulators and VR devices, the results would provide quantitative evidence that simulator sickness and VR sickness represent distinct forms of motion sickness. The theoretical implications of such a finding would be that the differences in sickness symptoms are driven by differences in the technological factors of the two types of VE systems, which would suggest that different technological solutions would be required to minimize side effects.

By identifying the relationship between the engineering characteristics of different VE systems and specific types of VE sickness symptoms, the research results will also provide VE system designers and engineers with a valuable tool that could be used to guide and direct their design efforts. For existing VE systems, the results of the research will provide information that can facilitate the identification of engineering modifications that should be implemented in a system which is producing a high incidence of sickness by making system designers, engineers, and evaluators aware of the system characteristics that contribute to specific types of symptoms. Similarly, an understanding of the major design features that affect sickness (and those which do not) can be used to direct the design and development of future VE technology. If system designers and engineers are aware of the equipment parameters that affect sickness, they can more readily determine which system features must be targeted for technology improvement in order to mitigate their impact on sickness. Moreover, for users of such systems, a focus on specific symptoms may lead to recommendations for different approaches to countermeasures for symptoms. In other words, if a device produced drowsiness, the treatment (i.e., the countermeasure) would be different than if the device produced balance problems.

One of the goals of the research was to determine the form of the relationship between different engineering features and the SSQ symptom subscales for different types of VE systems and to evaluate whether there was generalizability of this relationship over different VE devices. If the results of the research indicate that the relationships between key system variables which influence sickness are generalizable across different system configurations, the symptom profiles could then be used as a prescriptive tool to characterize and evaluate system differences. One of the main outcomes of the research is the identification of the equipment design features of VE systems that influence specific types of sickness symptoms (i.e., SSQ subscales). This will serve

to not only identify critical design variables, it will also provide testable hypotheses for different combinations of equipment features that can be evaluated in future research applications. With this knowledge, one could also predict the specific types of symptoms (i.e., the distribution of SSQ symptom subscales) that users would experience as a result of exposure to a particular VE system configuration. Accordingly, the results of this research could be used to support those that use VE systems by determining, a priori, the types of symptoms that may occur from the use of their system. The expected SSQ symptom profile could then be used to determine specific VE usage protocols and aid in the selection of appropriate post-exposure countermeasures to facilitate readaptation to the "real" world.

As VR and simulator technology continue to develop, it is anticipated that VE systems will become less expensive and thus, more widely accessible to diverse populations. The number of people that could experience adverse side effects will also increase resulting in a greater risk for product liability claims. Kennedy, Kennedy, and Bartlett (2002) emphasized the need for manufacturers and owners of VE systems to take proactive steps in order to minimize their legal liability. The authors outlined a seven-step system safety approach that could be used to assess the potential risks associated with the aftereffects of VE exposure to circumvent product liability issues. In the general order of application preference, the steps of their safety approach were design, remove, guard, warn, train, certify, and monitor and debrief. Knowledge of the equipment features that influence specific types of symptoms will provide a means for manufacturers and VE system owners to directly address four of the higher priority safety steps.

First, Kennedy, Kennedy, and Bartlett (2002) stated that products should be designed to minimize harm (i.e., eliminate hazards) to the user. VE system developers could use the results of the research to determine which design features should be replaced or modified in future

systems in order to eliminate, or at least reduce, the potential for sickness. Their second proactive step suggested that hazards should be removed from existing systems. An understanding of the potential impact that specific system characteristics have on different types of symptoms will allow engineers and system designers to identify equipment features in existing systems that contribute to the adverse effects and therefore, need to be modified. The authors also indicated that users need to be warned of any remaining hazards. By providing the ability to predict symptom patterns based on a specific system configuration, warnings about particular side effects that may be experienced during or after exposure can be developed for potential users. Finally, Kennedy, Kennedy, and Bartlett (2002) noted that the expected level of hazard imposed by a VE system should be determined (i.e., system certification). After identification of a specific threshold value for a system to be considered acceptable in terms of the degree of disturbance produced by human-VE interaction, companies could use the expected SSQ symptom profile produced by their system to certify the effectiveness of the system's freedom from hazard to the user.

CHAPTER THREE: METHODOLOGY

The objective of the current research was to identify an underlying symptom structure (i.e., SSQ profile) for different types of VE systems and then determine whether there were quantitative differences in the patterns of symptoms over diverse systems. Another goal of the research was to determine the form of the relationship between different engineering features and the SSQ symptom subscales for different types of VE systems (i.e., simulators and VR devices) and to evaluate whether there was generalizability of this relationship over different VE systems. Additionally, the terms cybersickness or virtual reality sickness are commonly used in the VE sickness literature to refer to the adverse effects produced by VR devices in order to distinguish the symptoms from those produced by simulators. Therefore, a second objective of the research was to determine whether the sickness produced by exposure to simulators and VR devices were quantitatively different. As stated previously, there is no consistent use and/or meaning of the term "sickness" in the scientific literature. Therefore, it is important to emphasize that "sickness" is used in the following sections to refer to the signs and symptoms of motion sickness that have been measured and scored on a standardized questionnaire (i.e., the SSQ).

Sickness Database

Over the past 20 years, Dr. Robert Kennedy and his colleagues have used the SSQ to collect data on motion sickness-like symptoms associated with exposure to various provocative environments including simulator and VR devices. Other scientists within the U.S. and abroad

have also used the SSQ in their research on motion sickness or to evaluate their study participants' physical condition after exposure to various systems under investigation (e.g., driving simulators). Several of these scientists provided their SSQ data to Dr. Kennedy for inclusion in the database. Access to the additional data not only served to increase the size of the SSQ database, but also increased the number of different devices represented in the database. The SSQ database contained approximately 13,500 pre/post exposure SSQ scores. The environments represented by the different datasets contained in the database included exposure to real motion stimuli (e.g., sea sickness, space sickness), simulated inertial motion stimuli, and visually-induced motion stimuli.

A subset of the SSQ data (i.e., simulator and VR systems) was used for this project. Specifically, the analyses for this project focused on the following five types of simulated environment (VE) systems: Military flight simulators (e.g., moving-base, fixed-base, CRT-display, Dome-display, etc.), driving simulators, and the three different VR display systems (i.e., HMD, BOOM, and CAVE). The data for simulator and VR systems in the SSQ database included pre/post exposure SSQ scores for approximately 3,745 participants. These data, however, only represented one exposure to a given device. If multiple exposure data were included, which will be discussed in a subsequent section, then the size of the database increased to over 5,200 pre/post exposure SSQ scores. The datasets that were available for the analyses included: 32 flight simulator studies, four driving simulator studies, 18 HMD studies, five Boom studies, and two CAVE studies. Therefore, the incidence and severity of VE sickness was evaluated across a broad spectrum of simulators (flight and driving) and VR devices. The equipment characteristics represented in the database for simulators included: simulator type (fixed-wing, rotary-wing, driving), platform (fixed-base, motion-base), the degrees of freedom of

the motion base, display type (CRT, Dome, Projection Screen), image generator (video camera, CGI), field-of-view, resolution, and the average system latency. The VR system characteristics included: display type (HMD, CAVE, Boom), display manufacturer and model (e.g., Virtual Research VR-6, Virtual i*O i*glasses!), display size, HMD visual display type (binocular, monocular), field-of-view, resolution, display weight, adjustability of the interpupillary distance (IPD), and the model, speed, and latency of the head tracker. Summary information for the database of the simulators and VR systems, including some of the characteristics of the equipment, are shown in Appendix B and Appendix C, respectively.

Quantification of System Profiles

Several different analyses were required in order to identify an underlying symptom structure (i.e., SSQ profile) for different types of VE systems and to quantitatively evaluate the differences in the SSQ profiles over diverse systems. Information regarding the specific analyses conducted on the SSQ data are presented in the following sections.

Database Organization

Before any analyses were performed, the SSQ datasets had to be organized. First, the data within each individual study in the SSQ database were inspected and cases with any missing post-exposure SSQ data were removed from the dataset. Then, the data within each study were coded with a "study number". Many of the flight simulator studies and a few of the VR studies contained data on individuals that received multiple exposures to the same device as well as individuals that only received one exposure. Therefore, the individual SSQ data within each study were first grouped by subject identification number, then the data within each dataset were

arranged into multiple and single exposure groups. Finally, each study was sorted and assigned a "system type" according to the type of system used in the study (e.g., fixed-wing flight simulator, rotary-wing flight simulator, driving simulator, HMD, BOOM, or CAVE).

Profile Development

Once the datasets were organized, an initial data "screening" analysis was conducted in order to determine the characteristics of the data that could be used in the subsequent analyses. For example, the literature review mentioned that the number of symptoms experienced from exposure to a simulated environment can vary; some people exhibit all or several of the symptoms while others exposed to the same device may only experience a few symptoms or no symptoms at all (Kennedy & Fowlkes, 1992). Because the focus of the research was on the type and severity of sickness produced by exposure to simulators and VR devices, only individuals that reported any type of symptoms after exposure to the VE system were included in the database (i.e., individuals with a Total Severity score of zero were eliminated). Also, many of the studies in the SSQ database contained two types of data, that is data for individuals with multiple exposures to the same device and individuals with only a single exposure. Consequently, another issue that was addressed was whether the analyses would be based on all of the SSQ data (i.e., include multiple exposures) or only a single exposure for each participant. Preliminary analyses on the datasets with multiple exposure data indicated that scores on the SSQ subscales and the resultant profile for single and multiple exposure data were fairly consistent. In order to make the data comparable for each study, however, only the first exposure data for individuals with multiple exposures was used for all subsequent testing so that each individual was represented once and only once in each dataset.

Profile Analyses

As discussed in the literature review, exposure duration is one of the major factors influencing VE sickness. Kennedy, Stanney, and Dunlap (2000) also stated "it will not be possible to perform quantitative meta-analytic comparisons of the variance accounted for by the disparate determiners of sickness unless time is taken into account" (p. 464). Consequently, this issue had to be addressed before any statistical analyses were conducted. First, it is important to note that their article only dealt with the issue of overall sickness severity (i.e., the SSQ TS scores), not the SSQ subscale scores. Moreover, a review of the scientific literature and the preliminary profile analysis investigations conducted by Kennedy and his colleagues suggested that duration would affect the *severity* of sickness (i.e., the level of the subscale scores), but may not affect the overall shape of the symptom profile within a given VE device. Accordingly, the data from a VR experiment which used the same VR system and visual stimulus, but varied exposure duration (15, 30, and 45 min.) between groups were examined. The SSQ profile data for the three exposure groups are presented in Figure 2 below.

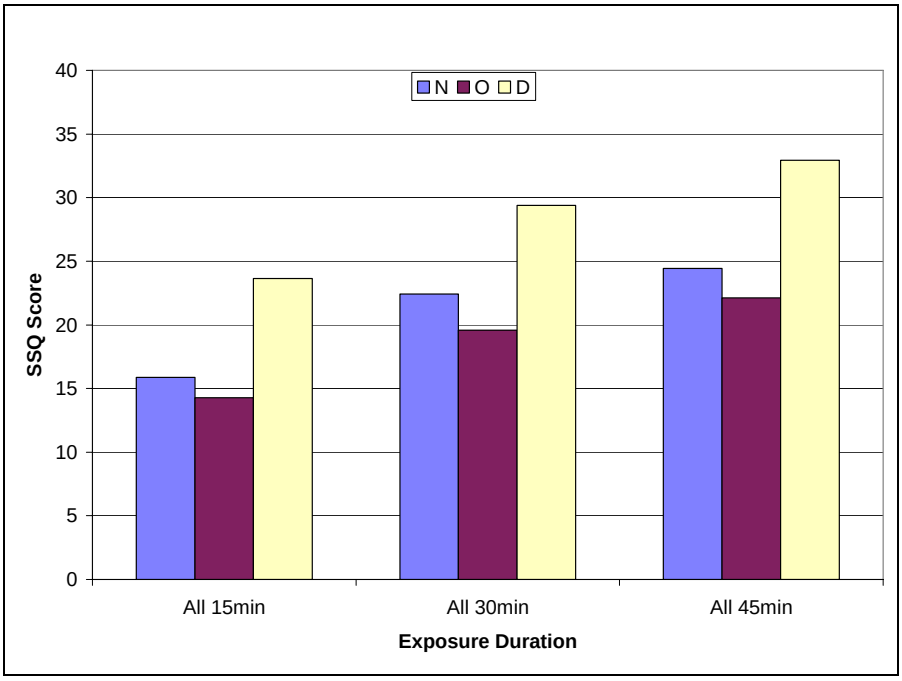

Figure 2. SSQ profiles for three different exposure duration groups.

The data in Figure 2 suggest that as the duration of exposure increases, the severity of symptoms also increase. However, the SSQ profile appears consistent across duration groups. These data provide some evidence that the relative contribution of the SSQ subscales are fairly insensitive to different exposure durations. Moreover, one of the goals of the research was to determine the relationship between the engineering features of the systems and the subscale scores. Therefore, exposure duration was not included as a variable in this research.

Analysis of Profile Differences Between and Within System Types

In order to test for differences among the profiles for each of the individual studies, each subscale score was converted to a proportion of the sum of the three subscale scores so that two types of Analysis of Variance (ANOVA) tests with follow-up multiple comparisons could be

used to detect relative differences in profiles among individual studies. Significant differences would indicate differences in profiles both within and between system types. Because use of the subscale scores would not capture the relative contribution of the subscales in the profile information, the subscale scores were adjusted so that profile differences were reflected. The adjustment used a proportional subscale score denoted:

a_i $\quad$ $i = 1, 2, 3$

where a_i = proportion of subscale i relative to the sum of the three subscale scores.

Using a_i for each individual "normalized" the scores so that only the relative positions of the subscale scores were considered. Several Multivariate ANOVAs (MANOVAs) were run in order to test all three of the proportional subscale scores simultaneously. First, a MANOVA was conducted in order to determine if there were profile differences between VE types (i.e., simulator and VR systems). Then a MANOVA was conducted to determine whether there were profile differences between the three types of simulators (i.e., Fixed-Wing, Rotary-Wing, and Driving simulators). Likewise, another MANOVA was conducted to determine whether there were profile differences between the three types of VR systems (i.e., HMD, BOOM, and CAVE). For each of these analyses, a significant difference indicated a different SSQ profile and was followed up with multiple comparison tests. The results of the analyses would reveal whether or not there were profile differences within and between system types. As an example of the type of results, for which specifics will be presented later, see Figure 3 which shows the SSQ profile for three hypothetical systems (a, b, c). While the profiles in Figure 3 are similar in that D>N>O, the relative contribution of the subscale factors are different for system "a" compared to system "b" and "c", which have similar relative subscale contributions. Using proportional subscale scores would allow detection of the difference between "a" and the other systems but

would not indicate differences between "a" and "b" even though the sum of the three subscale scores for "a" and "b" are different.

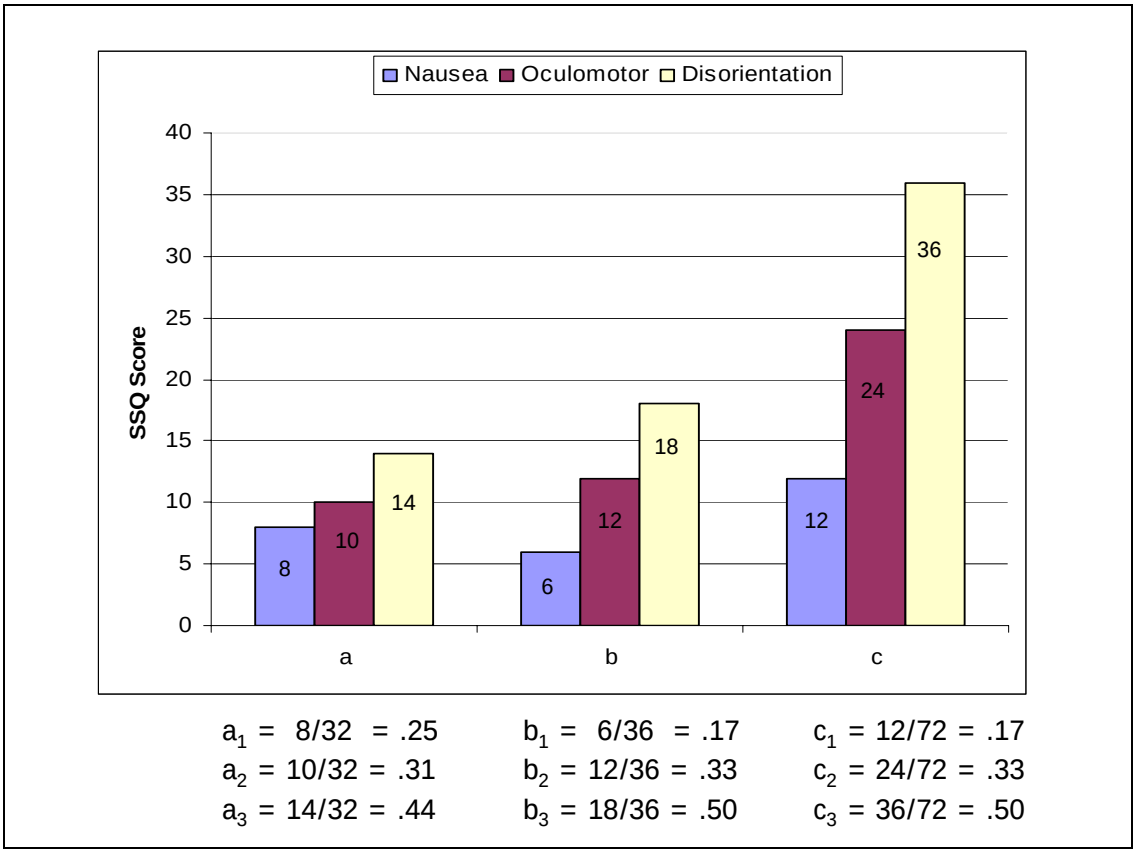

Figure 3. Similar SSQ profiles with different proportional variable subscales.

Next, two MANOVAs were run on the proportional variable subscale data from the individual studies (one each for simulator and VR systems) in order to evaluate profile differences among the individual studies. For both analyses, a significant difference indicated a different profile and was followed up with multiple comparison tests. A close examination of the similarities and differences in the subscales that were revealed in the analyses allowed comparisons between and within VE system types and were used to investigate design features that contributed to equal relative subscale scores (i.e., to identify a common cause). For

example, as shown in Figure 4, the profiles for two other hypothetical systems appear different from each other. However, an examination of the proportional variable subscale scores in Figure 4 indicate that both systems have the same relative contribution on the Disorientation subscale. The analyses, therefore, identified how the systems were similar (e.g., high Disorientation) and how they were different (e.g., different contributions of the Nausea and Oculomotor factors).

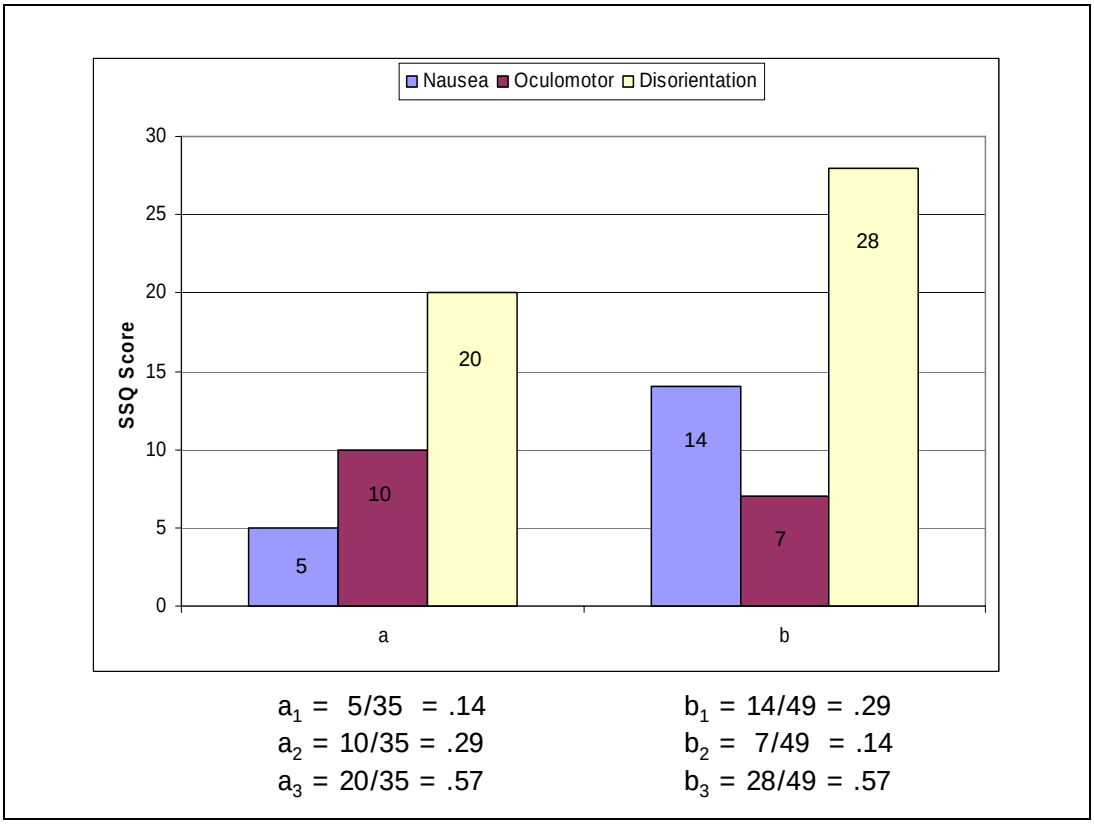

Figure 4. Different SSQ profiles with a similar proportional variable subscale.

Analysis of Differences in Sickness Severity Between and Within System Types

Two additional sets of analyses were performed on the actual SSQ data (i.e., not the proportional variables) in order to evaluate differences in sickness severity between and within VE system types as well as differences among the individual studies. The first set of analyses evaluated the SSQ Total Severity (TS) score to determine whether there were statistically

significant differences in overall sickness severity among the studies. First, a t-test was run to assess whether there were differences simulators and VR systems. A One-Way ANOVA was then run to determine whether there were differences among the three types of simulators (Fixed-Wing, Rotary-Wing, and Driving simulators). Another One-Way ANOVA was then run to determine whether there were differences among the various simulator studies. Next, a One-Way ANOVA was conducted to determine whether there were differences between the three types of VR devices (HMD, BOOM, CAVE). Lastly, a One-Way ANOVA was run to determine whether there were differences among the various VR studies. For all of the ANOVAs, a significant difference indicated a different TS score and was followed up with multiple comparison tests.

The second set of analyses evaluated scores on the individual SSQ subscales (Nausea, Oculomotor, and Disorientation). First, three One-Way ANOVAs were run on the subscale scores (one each for Nausea, Oculomotor, Disorientation) for the simulator studies in order to determine whether there were differences among the various studies. Similarly, three One-Way ANOVAs were run on the subscale scores for the VR studies in order to determine whether there were differences among the various VR studies. As with the Total Severity analyses, a significant difference indicated a different severity of sickness and was followed up with multiple comparison tests. Similarities and differences in sickness severity in the individual studies were then used to investigate design features that contributed to equal or different severities. Thus, a total of 16 different statistical analyses were conducted on the data; Table 2 provides a summary of the dependent variables and the specific hypotheses tested in each analysis. Attempts to explain all of the results in terms of a single unifying paradigm appear in Chapter 5.

Table 2. Dependent Variables and Research Hypotheses

Dependent Variables	**Hypothesis Tested**
Proportional Subscale Scores: - Nausea - Oculomotor - Disorientation Total Severity Score	H_o: Simulators = VRs H_o: Fixed-Wing = Rotary-Wing = Driving H_o: Simulator Study 1 = … = Simulator Study 21 H_o: HMD = BOOM = CAVE H_o: VR Study 1 = … = VR Study 16
Actual SSQ Subscale Scores: - Nausea - Oculomotor - Disorientation	H_o: Simulator Study 1 = … = Simulator Study 21 H_o: VR Study 1 = … = VR Study 16

Profile Validation

Once the relationship between the SSQ profiles and the engineering characteristics of a device were identified, additional SSQ data were used to cross-validate the results. Specifically, two datasets (one from a simulator and one from a VR system) that were *not* included in the original profile analyses were used to validate the conclusions that were derived from the preceding analyses.

CHAPTER FOUR: RESULTS

This chapter contains the results of the data analyses for the research conducted in order to identify similarities and differences both between and within different types of VE systems.

Sickness Database

Before any analyses were performed, several "screening" procedures were conducted on the database in order to eliminate individual SSQ data (i.e., individual cases) based on different exclusion criteria. First, all of the data were inspected and cases with any missing post-exposure SSQ data were removed from the dataset. The data within each individual study in the SSQ database that contained multiple exposure data were then reviewed and only data for the first exposure was retained (i.e., data for all subsequent exposures were eliminated from the dataset). Additionally, because the literature on motion sickness indicates that an individual's current physiological state can influence their susceptibility to motion sickness, any cases where the pre-exposure Total Severity scores were greater than 12.0 were eliminated from the database. Then, all cases where an individual did not report any symptoms after exposure to the VE system (i.e., the Total Severity score was zero) were eliminated. Next, the data for two studies (one simulator and one VR study) that would be used for the validation study were removed from the database. These datasets were chosen arbitrarily by reviewing the number of study participants in each of the simulator and VR datasets and selecting a study that appeared to have a sufficient number of cases. In the final phase of the database preparation. Finally, each study was evaluated and any

study that had an insufficient number of cases as well as those that did not have enough information on the various equipment features for the system used in the study were eliminated from further consideration. The final database that was used for the subsequent analyses, shown in Table 3, contained the following types of VE studies: eight Fixed-Wing flight simulator, nine Rotary-Wing flight simulator, four Driving simulator, 13 HMD, two BOOM, and one CAVE with a total of 2100 individuals. A list of references for the simulator and VR studies included in the final database is provided in Appendix D; Appendix E contains a list of references for additional studies that were available, but not included in the database.

Table 3. VE Studies and Number of Participants Used in the Statistical Analyses

	Type of VE System	Study Number	n
Simulators	Driving	201	62
		202	53
		203	43
		204	104
	Fixed-Wing	302	28
		303	18
		304	8
		306	10
		307	39
		308	20
		316	19
		318	8
	Rotary-Wing	305	86
		309	66
		310	67
		311	42
		312	125
		313	38
		314	30
		315	28
		317	14
VR Systems	HMD	101	47
		102	13
		103	25
		104	19
		105	81
		106	30
		107	200
		108	197
		109	194
		110	211
		111	32
		112	39
		113	12
	BOOM	650	25
		651	32
	CAVE	725	35
Total			2100

Analysis Methods

The methods chosen for the MANOVA and post hoc analyses are presented in the following two sections.

MANOVA Analyses

The information on the database used for the analyses shows that the number of participants varied among the different studies. Consequently, Pillai's Trace was used as the multivariate test statistic because it is considered to be a more appropriate test when there are small or unequal sample sizes.

Post Hoc Multiple Comparison Tests

Typically researchers use an alpha level of .05 or .01 to decide the significance of their results (i.e., whether to reject the null hypothesis). However, the purpose of the post-hoc testing on the individual simulator and VR studies was to explore which equipment parameters affect sickness that occurs as a result of exposure to VE systems (i.e., there were no a priori hypotheses). Thus, the goal of this phase of the analyses was hypothesis gathering as opposed to the more traditional hypothesis testing. In exploratory research, higher significance levels (e.g., .20) are generally used in order to avoid the possibility of overlooking potentially important data, which could occur when a conservative significance level is used (Cosby, 1993). Consequently, more liberal alpha levels (.10 and .15) were initially considered for two of the MANOVA analyses in order to determine whether the use of a higher significance level would provide more information on the post-hoc comparison tests. The results on the post hoc analyses for the VR

MANOVA indicated that no additional information was provided at α = .10 and only minor changes in one of the results for subscales occurred at α =.15. In contrast, the results on the post hoc analyses for the simulator MANOVA revealed that changes on all of the subscales occurred at α = .10, but more information on the significant differences between studies was available at α =.15. Therefore, a more conservative significance level (α =.05) was used for all of the post hoc analyses on the VR studies whereas the more liberal significance level (α =.15) was used for all of the post hoc analyses on the simulator studies.

Quantification of System Profiles

Several different analyses were required in order to identify an underlying symptom structure (i.e., SSQ profile) for different types of VE systems and to quantitatively evaluate the differences in the SSQ profiles over diverse systems. The first group of analyses involved tests on the proportional subscale scores in order to evaluate differences in profiles between and within VE system types. In contrast, the second group of analyses tested the actual SSQ Total Severity and subscale scores in order to evaluate differences in sickness severity both within and between VE system types. Information regarding the specific analyses conducted on the SSQ data are presented in the subsequent sections.

Analysis of Profile Differences Between and Within System Types

The results for the MANOVAs conducted on the proportional subscales scores and the follow-up multiple comparison tests are presented in the following sections.

MANOVA on Profile Differences Between VE System Type

A multivariate analysis of variance (MANOVA) was conducted to assess if there were profile differences between the two types of virtual environment (VE) systems (simulators and VR devices). A statistically significant difference was found between the two types of VE systems, Pillai's Trace = .069, $F(1, 2097) = 78.08$, $p < .001$. The means and standard deviations of the proportional subscale scores for the two types of VE systems are shown in Table 4.

Table 4. Means and Standard Deviations Comparing Simulators and VR Devices

		Proportional N		Proportional O		Proportional D	
VE System Type	n	M	SD	M	SD	M	SD
Simulators	908	0.297	0.293	0.497	0.334	0.206	0.247
VR Devices	1192	0.273	0.247	0.380	0.287	0.347	0.268
Total	2100	0.283	0.268	0.431	0.313	0.286	0.269

Follow up Univariate ANOVAs indicated that all three of the proportional subscale scores were significantly different between the two types of VE systems: $F(1, 2098) = 4.06$, $p = .044$ for the proportional Nausea subscale score; $F(1, 2098) = 74.68$, $p < .001$ for the proportional Oculomotor subscale score; and $F(1, 2098) = 152.24$, $p < .001$ for the proportional Disorientation subscale score. As shown in Table 4 above, the mean for Simulators was greater than VR Devices for the proportional Nausea and Oculomotor subscale scores, whereas the mean for VR Devices was greater for the proportional Disorientation subscale score.

MANOVA on Profiles for Simulator System Type

A MANOVA was conducted to assess if there were differences among the three types of simulators (i.e., Fixed- and Rotary-Wing flight simulators and Driving simulators). A statistically significant difference in profiles was found among the different types of simulators, Pillai's Trace = .099, $F(4, 1810) = 23.44$, $p < .001$. The means and standard deviations of the proportional subscale scores for the three types of simulators are shown in Table 5.

Table 5. Means and Standard Deviations Comparing Three Types of Simulators on the Proportional Subscale Scores

		Proportional N		Proportional O		Proportional D	
Simulator Type	n	M	SD	M	SD	M	SD
Fixed-Wing	150	0.304	0.349	0.585	0.362	0.111	0.195
Rotary-Wing	496	0.293	0.291	0.532	0.332	0.175	0.234
Driving	262	0.301	0.260	0.382	0.290	0.318	0.259
Total	908	0.297	0.293	0.497	0.334	0.206	0.247

Levene's test indicated that the error variances were not equal and therefore, follow up multiple comparisons were conducted using the Games-Howell test. These analyses indicated that the means for the three types of simulators did not differ significantly on the proportional Nausea subscale score. In contrast, Driving simulators were significantly different than Fixed-Wing ($p < .001$) and Rotary-Wing ($p < .001$) flight simulators on the proportional Oculomotor subscale score. The means presented previously in Table 5 show that Fixed-Wing (.585) and Rotary-Wing (.532) simulators produce larger proportional Oculomotor scores than Driving simulators (.382). However, the post hoc tests revealed that scores on this subscale for Fixed-Wing and Rotary-Wing simulators were not significantly different.

The post hoc analyses on the proportional Disorientation subscale score revealed that there were statistically significant mean differences between each of the three types of simulators. Specifically, the Fixed-Wing and Rotary-Wing simulators were significantly different than Driving simulators, $p < .001$ for both comparisons and the Fixed- and Rotary-Wing simulators were significantly different from each other ($p = .002$). As shown in Table 5 above, Driving simulators produce the largest mean proportional Disorientation score (.318) followed by Rotary-Wing simulators (.175) and Fixed-Wing simulators (.111).

MANOVA on Profiles Among Simulator Studies

In order to determine whether there were differences between various types of Simulator system configurations, a MANOVA was conducted on the 21 simulator studies, where each study represented a homogenous set of equipment features. The results of the analysis revealed a statistically significant difference among the discrete simulator studies, Pillai's Trace = .199, $F(40, 1774) = 4.89$, $p < .001$. Table 6 presents the descriptive statistics (i.e., means and standard deviations) of the proportional subscale scores for all of the simulator studies, which are grouped by type of simulator system.

Table 6. Means and Standard Deviations Comparing Studies Using Three Types of Simulator Systems

Simulator Type	Study Number	n	Proportional N		Proportional O		Proportional D	
			M	SD	M	SD	M	SD
Driving	201	62	0.322	0.239	0.310	0.210	0.368	0.256
	202	53	0.261	0.185	0.336	0.179	0.403	0.181
	203	43	0.267	0.269	0.369	0.304	0.364	0.299
	204	104	0.322	0.296	0.453	0.352	0.225	0.252
Fixed-Wing	302	28	0.459	0.399	0.402	0.385	0.140	0.228
	303	18	0.277	0.215	0.515	0.207	0.208	0.206
	304	8	0.308	0.375	0.628	0.424	0.064	0.181
	306	10	0.246	0.332	0.679	0.374	0.075	0.162
	307	39	0.157	0.260	0.756	0.290	0.087	0.176
	308	20	0.331	0.285	0.589	0.321	0.080	0.201
	316	19	0.451	0.465	0.463	0.437	0.087	0.155
	318	8	0.195	0.379	0.677	0.385	0.128	0.248
Rotary-Wing	305	86	0.217	0.249	0.581	0.334	0.202	0.295
	309	66	0.384	0.355	0.500	0.373	0.116	0.196
	310	67	0.264	0.271	0.566	0.334	0.170	0.226
	311	42	0.320	0.278	0.474	0.288	0.206	0.247
	312	125	0.329	0.282	0.465	0.298	0.206	0.211
	313	38	0.287	0.250	0.523	0.313	0.189	0.249
	314	30	0.206	0.213	0.649	0.303	0.146	0.219
	315	28	0.341	0.405	0.581	0.424	0.078	0.155
	317	14	0.166	0.232	0.659	0.324	0.176	0.242
	Total	908	0.297	0.293	0.497	0.334	0.206	0.247

The assumption of equal error variances was violated for all of the proportional subscale scores (i.e., Levene's test was significant) and therefore, the Games-Howell test was used for the follow up multiple comparisons. A summary of the post hoc test results for the proportional Nausea subscale scores is presented first in Table 7. In each of these tables, the data for the simulator studies is ordered according to the magnitude of the mean score, from highest to lowest, along the horizontal and vertical axes. Within each table, a significant difference (α =

.15) between two studies is indicated with an asterisk (*) in the cell corresponding to the intersection of the two studies, whereas the cells for non-significant study pairs are empty. It is important to note however, that significant differences between studies are only represented in the cells above the diagonal of the table since the cells below the table's diagonal are simply a mirror image of those above it. Thus, the table is arrayed similar to the data in a correlation matrix.

Table 7. Results of the Multiple Comparison Tests on the Proportional Nausea Subscale Scores for Simulator Studies

	Study	302	316	309	315	308	312	201	204	311	304	313	303	203	310	202	306	305	314	318	317	307
Study	Mean	.459	.451	.384	.341	.331	.329	.322	.322	.320	.309	.287	.277	.267	.264	.261	.246	.217	.206	.195	.166	.157
302	0.459																					*
316	0.451																					
309	0.384																	*				*
315	0.341																					
308	0.331																					
312	0.329																					*
201	0.322																					
204	0.322																					
311	0.320																					
304	0.308																					
313	0.287																					
303	0.277																					
203	0.267																					
310	0.264																					
202	0.261																					
306	0.246																					
305	0.217																					
314	0.206																					
318	0.195																					
317	0.166																					
307	0.157																					

* Significant difference ($p < .15$)

The results of the post hoc test indicated that there were no significant differences on the proportional Nausea subscale for the comparisons within any of the studies that used a Driving simulator. There also were no significant differences for the comparisons between the Driving simulator studies and any of the flight simulator studies. However, the post hoc test results did reveal a significant difference in mean proportional Nausea scores between two Fixed-Wing studies (Study 302 and 307), two Rotary-Wing studies (Study 305 and 309), and two separate comparisons between a Fixed- and Rotary-Wing study (Study 307 and 309 and Study 307 and 312). The means for the Fixed-Wing, within simulator type comparison indicate that the simulator used in Study 302 was significantly higher than the simulator from Study 307 (cf. Table 7). The direction of the effect for the other within simulator type comparison is that the Rotary-Wing simulator in Study 309 had a significantly higher mean score than the one used in Study 305. In both of the between flight simulator type comparisons, the Rotary-Wing simulators (Studies 309 and 312) showed significantly higher mean proportional Nausea scores than the Fixed-Wing simulator (Study 307).

The differences in equipment features between each pair of studies that were identified as statistically different on this subscale and the associated significance level for each comparison is presented in Table 8 below. Within the table, the study pairs are ordered from largest to smallest mean difference on the proportional Nausea subscale. However, caution is required when interpreting these results. While the difference in mean scores between studies may be attributable to differences in the equipment features of the systems used for each study pair, it is also possible that the differences could merely be an artifact of the error rate for the test. Specifically, based on the significance level used for the post hoc test ($\alpha = .15$), there is a 15% probability (i.e., 32 out of the 210 comparisons) that any of the significant findings may have

occurred by chance; this analysis only revealed four comparisons that were significantly different.

Table 8. Equipment Differences and Significance Levels for Simulator Studies with Significantly Different Proportional Nausea Scores

Study Pair	Simulator Type	Mean Difference	Significance	Equipment Differences[1]
302, 307	Fixed, Fixed	0.302	$p = .103$	D, F, M
307, 309	Fixed, Rotary	0.228	$p = .036$	F, S
305, 309	Rotary, Rotary	0.167	$p = .146$	F
307, 312	Fixed, Rotary	0.173	$p = .078$	F, L, S

[1] D = Display Type, F = Field of View, L = System Latency, M = Motion Base, S = Simulator Type

A review of Table 8 reveals that all of the study comparisons which had significantly different proportional Nausea scores differed in terms of the display's field-of-view. The Fixed-Wing, within system type comparison also had a difference in the type of display and the motion base. Specifically, Study 302 had a Dome display without a motion base whereas Study 307 had a CRT display with a motion base. In contrast, Study 307 and 312, a between flight simulator type comparison, both systems had a CRT display and motion base, but differed in the overall system latency.

The results of the post hoc analysis on the proportional Oculomotor subscale scores for the Simulator studies are summarized in Table 9.

Table 9. Results of the Multiple Comparison Tests on the Proportional Oculomotor Subscale Scores for Simulator Studies

	Study	307	306	318	317	314	304	308	305	315	310	313	303	309	311	312	316	204	302	203	202	201
Study	Mean	.756	.679	.677	.659	.649	.628	.589	.581	.581	.566	.523	.515	.500	.474	.465	.463	.453	.402	.369	.336	.310
307	0.756											*	*	*	*	*		*	*	*	*	*
306	0.679																					
318	0.677																					
317	0.659																				*	*
314	0.649																			*	*	*
304	0.628																					
308	0.589																					*
305	0.581																			*	*	*
315	0.581																					
310	0.566																				*	*
313	0.523																				*	*
303	0.515																					*
309	0.500																					*
311	0.474																					
312	0.465																				*	*
316	0.463																					
204	0.453																					*
302	0.402																					
203	0.369																					
202	0.336																					
201	0.310																					

* Significant difference ($p < .15$)

The results of the post hoc analysis presented in Table 9 above indicated a significant difference in the mean scores of several Driving and Fixed-Wing study comparisons (n = 6 pairs) as well as 15 Driving and Rotary-Wing study pairs. In all of these significant Driving and Flight simulator (Fixed- and Rotary-Wing) comparisons, the mean proportional Oculomotor scores for Driving simulators were lower than either type of flight simulator. Four of the Fixed- and Rotary-Wing study comparisons were significantly different. In each of these comparisons, the mean for the Fixed-Wing study (Study 307) was significantly greater than the Rotary-Wing mean (Studies 309, 311, 312, and 313). The results in Table 9 also reveal a significant difference in means for some of the within simulator type comparisons. Specifically, in the Driving simulator study pair, the mean proportional Oculomotor score was significantly greater for Study 204 than Study 201. The Fixed-Wing comparisons showed two significant study pairs in which the simulator in Study 307 had a greater mean score than the simulators in Studies 302 and 303. However, the results indicated that there were no significant differences for the Rotary-Wing, within simulator type comparisons.

Table 10 provides a list of the equipment features that differed between each pair of studies identified as statistically different on this subscale, as well as the associated significance level for each comparison. The data within the table are grouped according to the type of simulator pair and within each group, the data ordered from largest to smallest mean difference. Again, caution is required when interpreting these results. Only 28 comparisons showed significantly different mean scores and, based on the error rate, 32 out of the 120 comparisons could have occurred merely by chance.

Table 10. Equipment Differences and Significance Levels for Simulator Studies with Significantly Different Mean Proportional Oculomotor Scores

Study Pair	Simulator Type	Mean Difference	Significance	Equipment Differences[1]
201, 307	Driving, Fixed	0.4465	$p < .001$	D, F, L, M, S
202, 307	Driving, Fixed	0.4202	$p < .001$	F, L, M, S
203, 307	Driving, Fixed	0.3869	$p < .001$	D, F, L, S
204, 307	Driving, Fixed	0.3032	$p < .001$	D, F, MD, S
201, 308	Driving, Fixed	0.2790	$p = .100$	D, F, M, S
201, 303	Driving, Fixed	0.2055	$p = .083$	D, F, M, S
201, 317	Driving, Rotary	0.3487	$p = .094$	D, F, M, S
201, 314	Driving, Rotary	0.3388	$p < .001$	D, F, M, S
202, 317	Driving, Rotary	0.3224	$p = .149$	F, M, S
202, 314	Driving, Rotary	0.3125	$p = .001$	F, M, S
203, 314	Driving, Rotary	0.2792	$p = .034$	D, F, S
201, 305	Driving, Rotary	0.2714	$p < .001$	D, F, L, M, S
201, 310	Driving, Rotary	0.2564	$p < .001$	D, F, L, M, S
202, 305	Driving, Rotary	0.2451	$p < .001$	F, L, M, S
202, 310	Driving, Rotary	0.2300	$p = .001$	F, L, M, S
201, 313	Driving, Rotary	0.2135	$p = .053$	D, F, M, S
203, 305	Driving, Rotary	0.2118	$p = .061$	D, F, L, S
201, 309	Driving, Rotary	0.1902	$p = .063$	D, F, M, S
202, 313	Driving, Rotary	0.1871	$p = .147$	F, M, S
201, 312	Driving, Rotary	0.1551	$p = .010$	D, F, L, M, S
202, 312	Driving, Rotary	0.1287	$p = .064$	F, L, M, S
201, 204	Driving, Driving	0.1433	$p = .132$	F, M, R
302, 307	Fixed, Fixed	0.3546	$p = .020$	D, F, M
303, 307	Fixed, Fixed	0.2411	$p = .085$	F
307, 312	Fixed, Rotary	0.2915	$p < .001$	F, L, S
307, 311	Fixed, Rotary	0.2824	$p = .006$	F, L, S
307, 309	Fixed, Rotary	0.2563	$p = .023$	F, S
307, 313	Fixed, Rotary	0.2331	$p = .117$	S

[1] D = Display Type, F = Field of View, L = System Latency, M = Motion Base, MD = Motion Base DOF, R = Resolution (driving sims only), S = Simulator Type

The organization of the data in Table 10 provides an easily visible difference in the type of simulator, which is present in 89% of the significant study comparisons. The equipment features shown in Table 10 for the significant study comparisons also indicate that the

differences related to the display's field-of-view were present in all of the significant study pairs, except one (Study pair 307 and 313; a Fixed- and Rotary-Wing comparison).

Whether motion was provided by the simulator (i.e., motion versus no-motion) was another equipment difference that was present in a majority of the significant study comparisons. Specifically, a difference in the motion base was noted in 67% of the Driving and Fixed-Wing comparisons, 87% of the Driving and Rotary-Wing comparisons, the Driving-Driving simulator study pair, and in one of the two significant Fixed-Wing, within system type study comparisons. Notably, motion base differences were not present in any of the significant Fixed- and Rotary-Wing comparisons. Relatedly, overall system latency (i.e., the time between operator input to the system and those changes reflected in the visual display and motion base) was an equipment feature that differed in 50% of the Driving and Fixed-Wing and Fixed- and Rotary-Wing comparisons as well as in 47% of the Driving and Rotary-Wing study pairs.

The other major equipment difference, shown in a little more than half of the significant comparisons in Table 10, was the type of display. Differences in the type of display occurred in 83% of the Driving and Fixed-Wing study pairs and 60% of the Driving and Rotary-Wing comparisons. However, display differences were not present in any of the significant Fixed- and Rotary-Wing nor in the only within system Driving simulator comparison.

Lastly, Table 11 provides a results summary for the post hoc analysis on the proportional Disorientation subscale scores for the VR studies.

Table 11. Results of the Multiple Comparison Tests on the Proportional Disorientation Subscale Scores for Simulator Studies

	Study	202	201	203	204	303	311	312	305	313	317	310	314	302	318	309	316	307	308	315	306	304
Study	Mean	.403	.368	.364	.225	.208	.206	.206	.202	.189	.176	.170	.146	.140	.128	.116	.087	.087	.080	.078	.075	.064
202	0.403				*	*	*	*	*	*		*	*	*		*	*	*	*	*	*	*
201	0.368				*			*	*	*		*	*	*		*	*	*	*	*	*	*
203	0.364											*	*	*		*	*	*	*	*	*	*
204	0.225																	*		*		
303	0.208																					
311	0.206																					
312	0.206																	*		*		
305	0.202																					
313	0.189																					
317	0.176																					
310	0.170																					
314	0.146																					
302	0.140																					
318	0.128																					
309	0.116																					
316	0.087																					
307	0.087																					
308	0.080																					
315	0.078																					
306	0.075																					
304	0.064																					

* Significant difference ($p < .15$)

The results in Table 11 show that three of the within system type comparisons were significantly different. Inspection of the proportional Disorientation means for the Rotary-Wing study pair indicate that Study 312 was greater than Study 315 and the mean for Driving simulator Study 204 was significantly less than the means for Driving simulator Studies 201 and 202. None of the Fixed-Wing, within system type study comparisons were significantly different. Moreover, only one of the Fixed- and Rotary-Wing comparisons was significantly different. The means for these two simulators showed that Study 312 (Rotary-Wing) had a significantly greater mean than Study 307 (Fixed-Wing).

The majority of the significant differences in proportional Disorientation mean scores involved the Driving and Fixed-Wing and Driving and Rotary-Wing study comparisons. In most of these system comparisons, the means for the Driving simulators were significantly greater than the Fixed- (85% of the study pairs) and Rotary-Wing (95% of the study pairs) scores. The exceptions to this direction of effect were three Driving and Fixed-Wing comparisons where the mean proportional Disorientation scores for the Fixed-Wing studies (Studies 304, 306, and 308) were significantly greater than the mean for the Driving simulator in Study 201. The other exception was one of the significant Driving and Rotary-Wing comparisons which showed that the Rotary-Wing simulator (Study 315) had a greater mean than the Driving simulator used in Study 201.

Equipment features that differed between each pair of significantly different studies on the proportional Disorientation subscale and the associated significance level for each comparison is presented below in Table 12. The data within the table are grouped according to the type of simulator pair and within each group, the data ordered from largest to smallest mean difference. As with the proportional Nausea and Oculomotor analyses, caution is required when

interpreting these results due to the small number of significant comparisons relative to the error rate (i.e., 32 out of the 120 comparisons could have occurred merely by chance).

Table 12. Equipment Differences and Significance Levels for Simulator Studies with Significantly Different Mean Proportional Disorientation Scores

Study Pair	Simulator Type	Mean Difference	Significance	Equipment Differences[1]
202, 204	Driving, Driving	0.1777	$p < .001$	D, F, M, R
201, 204	Driving, Driving	0.1432	$p = .074$	F, M, R
202, 304	Driving, Fixed	0.3389	$p = .037$	D, F, S
202, 306	Driving, Fixed	0.3277	$p = .005$	F, S
202, 308	Driving, Fixed	0.3228	$p < .001$	F, M, S
202, 307	Driving, Fixed	0.3159	$p < .001$	F, L,M, S
202, 316	Driving, Fixed	0.3156	$p < .001$	D, F, M, S
201, 304	Driving, Fixed	0.3044	$p = .072$	D, F, S
203, 304	Driving, Fixed	0.2998	$p = .101$	F, M, S
201, 306	Driving, Fixed	0.2933	$p = .014$	D, F, S
203, 306	Driving, Fixed	0.2887	$p = .029$	D, F, M, S
201, 308	Driving, Fixed	0.2883	$p = .001$	D, F, M, S
203, 308	Driving, Fixed	0.2837	$p = .007$	D, F, S
201, 307	Driving, Fixed	0.2815	$p < .001$	D, F, M, S
201, 316	Driving, Fixed	0.2812	$p < .001$	D, F, L, M, S
203, 307	Driving, Fixed	0.2768	$p < .001$	D, F, L, S
203, 316	Driving, Fixed	0.2766	$p = .002$	F(V), S
203, 302	Driving, Fixed	0.2241	$p = .074$	F, M, S
201, 302	Driving, Fixed	0.2287	$p = .012$	D, F, S
202, 303	Driving, Fixed	0.1946	$p = .112$	F, M, S
202, 302	Driving, Fixed	0.1777	$p < .001$	D, F, S
204, 307	Driving, Fixed	0.1382	$p = .047$	D, F, MD, S
202, 315	Driving, Rotary	0.3224	$p < .001$	F, L, M, S
201, 315	Driving, Rotary	0.2900	$p < .001$	D, F, L, M, S
202, 309	Driving, Rotary	0.2871	$p < .001$	F, M, S
203, 315	Driving, Rotary	0.2854	$p < .001$	D, F, L, S
202, 314	Driving, Rotary	0.2569	$p < .001$	F, M, S
201, 309	Driving, Rotary	0.2527	$p < .001$	D, F, M, S
203, 309	Driving, Rotary	0.2480	$p = .002$	D, F, S
202, 310	Driving, Rotary	0.2325	$p < .001$	F, L, M, S
201, 314	Driving, Rotary	0.2225	$p = .008$	D, F, M, S
203, 314	Driving, Rotary	0.2178	$p = .069$	D, F, S

202, 313	Driving, Rotary	0.2133	$p = .005$	F, M, S
202, 305	Driving, Rotary	0.2009	$p < .001$	F, L, M, S
201, 310	Driving, Rotary	0.1981	$p = .002$	D, F, L, M, S
202, 312	Driving, Rotary	0.1970	$p < .001$	F, L, M, S
202, 311	Driving, Rotary	0.1966	$p = .008$	F, L, M, S
203, 310	Driving, Rotary	0.1935	$p = .062$	D, F, L, S
201, 312	Driving, Rotary	0.1625	$p = .006$	D, F, L, M, S
201, 305	Driving, Rotary	0.1665	$p = .047$	D, F, L, M, S
201, 313	Driving, Rotary	0.1788	$p = .098$	D, F, M, S
204, 315	Driving, Rotary	0.1468	$p = .035$	D, F, MD, S
307, 312	Fixed, Rotary	0.1189	$p = .085$	F, L, S
312, 315	Rotary, Rotary	0.1275	$p = .065$	F, L

[1] D = Display Type, F = Field of View, L = System Latency, M = Motion Base, MD = Motion Base DOF, R = Resolution (driving sims only), S = Simulator Type

A review of Table 12 indicates that a difference in the display's field-of-view was an equipment feature that occurred in all of the significant study comparisons. The arrangement of the data in Table 12 also makes it readily apparent that the type of simulator was a difference present in most (93%) of the significant comparisons. The presence or absence of simulated motion was another equipment difference noted in more than half (61%) of the significant comparisons. In particular, both of the significant comparisons between Driving simulator study pairs had a difference in whether motion was used, 75% of the Driving and Rotary-Wing comparisons showed a motion-base difference, and 50% of the Driving and Fixed-Wing study pairs. Additionally, overall system latency differed in almost 40% of the significant study comparisons. While only 20% of the Driving and Fixed-Wing study pairs had a difference in system latency, this difference was present in 55% of the Driving and Rotary-Wing comparisons as well as the significant Fixed- and Rotary-Wing study pair, and the within system, Rotary-Wing comparison.

Finally, differences in the type of display were present in a large number (61%) of the significant comparisons. Specifically, this feature differed in 65% of the Driving and Fixed-Wing study pairs, 60% of the Driving and Rotary-Wing pairs, and one of the two significant within system, Driving simulator comparisons. However, the displays were not different in the Fixed-and Rotary-Wing and the Rotary-Rotary Wing significant study pairs.

MANOVA on Proportional Subscale Scores for VR System Type

A MANOVA was conducted to assess if there were differences between the three types of VR systems (i.e., HMD, BOOM, and CAVE). A statistically significant difference was found among the different types of VR systems, Pillai's Trace = .028, $F(4, 2378) = 8.42$, $p < .001$. Table 13 presents the descriptive statistics (i.e., means and standard deviations) of each proportional subscale score for the three types of VR systems.

Table 13. Means and Standard Deviations Comparing Three Types of VR Systems

		Proportional N		Proportional O		Proportional D	
VR System	n	M	SD	M	SD	M	SD
HMD	1100	0.281	0.245	0.368	0.278	0.351	0.264
BOOM	57	0.217	0.232	0.511	0.294	0.272	0.272
CAVE	35	0.105	0.258	0.550	0.416	0.346	0.383
Total	1192	0.273	0.247	0.380	0.287	0.347	0.268

Levene's test indicated that the error variances of the proportional Nausea score were not statistically different. Accordingly, the Bonferroni test was used for the follow up multiple comparisons. This analysis showed that HMDs were significantly different than CAVE systems ($p < .001$). Inspection of the means presented in Table 13 above indicate that HMDs (.281)

produce a larger proportional Nausea score than CAVE systems (.105). Conversely, there were no significant differences detected in mean scores on this subscale for the comparisons between HMD and BOOM systems nor between BOOM and CAVE systems.

The post hoc analyses on the proportional Oculomotor and Disorientation subscale scores were conducted using the Games-Howell test because Levene's test of the equality of error variances was significant. The results for the proportional Oculomotor subscale revealed significant mean differences between the HMD and BOOM systems (p = .002) and between HMD and CAVE systems (p = .038). As shown in Table 13 above, the mean proportional Oculomotor score was greater for BOOM (.511) and CAVE (.550) systems than for HMDs (.368). However, the test on the proportional Oculomotor subscale means failed to reveal a significant difference between BOOM and CAVE systems. Similarly, the multiple comparison tests detected there were no significant differences on the proportional Disorientation subscale scores between any of the three types of VR systems.

MANOVA on Proportional Subscale Scores for VR Studies

In order to determine whether there were profile differences among various types of VR system configurations, a MANOVA was conducted on the 16 VR studies, where each study represented a homogenous set of equipment features. The results of the analysis revealed a statistically significant difference among the discrete VR studies, Pillai's Trace = .081, $F(30, 2352) = 3.32$, $p < .001$. Table 14 presents the descriptive statistics (i.e., means and standard deviations) of the proportional subscale scores for all of the VR studies, which are grouped by type of VR system.

Table 14. Means and Standard Deviations Comparing Studies Using Three Types of VR Systems

			Proportional N		Proportional O		Proportional D	
VR System	Study Number	n	M	SD	M	SD	M	SD
HMD	101	47	0.331	0.303	0.335	0.262	0.334	0.270
	102	13	0.336	0.175	0.322	0.224	0.342	0.198
	103	25	0.208	0.238	0.451	0.288	0.341	0.292
	104	19	0.331	0.357	0.410	0.378	0.259	0.298
	105	81	0.246	0.198	0.444	0.275	0.311	0.241
	106	30	0.241	0.232	0.483	0.320	0.276	0.269
	107	200	0.297	0.263	0.293	0.255	0.410	0.292
	108	197	0.293	0.227	0.356	0.269	0.350	0.259
	109	194	0.295	0.231	0.357	0.266	0.348	0.243
	110	211	0.278	0.201	0.392	0.256	0.330	0.219
	111	32	0.244	0.329	0.302	0.298	0.454	0.341
	112	39	0.213	0.363	0.409	0.364	0.378	0.341
	113	12	0.162	0.304	0.732	0.329	0.106	0.162
BOOM	650	25	0.228	0.221	0.509	0.333	0.263	0.257
	651	32	0.209	0.244	0.513	0.264	0.279	0.287
CAVE	725	35	0.105	0.258	0.550	0.416	0.346	0.383
	Total	1192	0.273	0.247	0.380	0.287	0.347	0.268

Since the assumption of equal error variances was violated (i.e., Levene's test was significant) for all of the proportional subscale scores, the Games-Howell test was used for the follow up multiple comparisons. A summary of the post hoc test results for the proportional Nausea subscale scores is presented first in Table 15. Similar to the post hoc results for the simulator studies, the data within each of the multiple comparison results tables for the simulator studies is ordered according to the magnitude of the mean score, from highest to lowest, along the horizontal and vertical axes. Also, a significant difference between two studies is indicated with an asterisk (*) in the cell corresponding to the intersection of the two studies, whereas the cells for non-significant study pairs are empty.

Table 15. Results of the Multiple Comparison Tests on the Proportional Nausea Subscale Scores for VR Studies

	Study	102	104	101	107	109	108	110	105	111	106	650	112	651	103	113	725
Study	Mean	0.336	0.331	0.331	0.297	0.295	0.293	0.278	0.246	0.244	0.241	0.228	0.213	0.209	0.208	0.162	0.105
102	0.3361																
104	0.3311																
101	0.3309																*
107	0.2970																*
109	0.2947																*
108	0.2934																*
110	0.2779																*
105	0.2455																
111	0.2440																
106	0.2411																
650	0.2275																
112	0.2130																
651	0.2088																
103	0.2084																
113	0.1615																
725	0.1049																

* Significant difference ($p < .05$)

As shown in Table 15 above, the results of the post hoc test indicated a significant difference only between the mean scores of several HMD studies and the CAVE study. In all of these comparisons, the mean proportional Nausea score was higher for the HMD studies than for the CAVE study. Table 15 also shows that there were no significant differences on this subscale for the comparisons between HMD and BOOM studies nor for the BOOM and CAVE study comparisons.

The differences in equipment features between each pair of studies that were identified as statistically different on this subscale and the associated significance level for each comparison is presented in Table 16 below. Within the table, the study pairs are ordered from largest to smallest mean difference on the proportional Nausea subscale. However, caution is required when interpreting these results. While the difference in mean scores between studies may be attributable to differences in the equipment features of the systems used in each study pair, it is also possible that the differences could merely be an artifact of the error rate for the test. Specifically, based on the significance level used for the post hoc test ($\alpha = .05$), there is a 5% probability (i.e., 6 out of the 120 comparisons) that any of the significant findings may have occurred by chance; this analysis only revealed five comparisons that were significantly different.

Table 16. Equipment Differences and Significance Levels for VR Studies with Significantly Different Proportional Nausea Scores

Study Pair	VR Type	Mean Difference	Significance	Equipment Differences[1]
101, 725	HMD, CAVE	0.2260	$p = .038$	D, F, TS, W
107, 725	HMD, CAVE	0.1920	$p = .015$	D, F, I, TS, W
109, 725	HMD, CAVE	0.1898	$p = .015$	D, F, I, TS, W
108, 725	HMD, CAVE	0.1884	$p = .016$	D, F, I, TS, W
110, 725	HMD, CAVE	0.1729	$p = .035$	D, F, I, TS, W

[1] D = Display Type, F = Field of View, I = IPD Adjust, TS = Head Tracker Speed, W = Display Weight

A review of Table 16 reveals that all of the study comparisons which had significantly different proportional Nausea scores differed in terms of the type of display, field-of-view, speed of the head tracker, and the weight of the display. Additionally, the ability to adjust the display's interpupillary distance (IPD) differed in each of the significant study comparisons (the HMD studies had an adjustable IPD whereas the CAVE system didn't) with the exception of the Study 101 and 725 comparison where both systems lacked an IPD adjustment feature.

The results of the post hoc analysis on the proportional Oculomotor subscale scores for the VR studies are summarized in Table 17.

Table 17. Results of the Multiple Comparison Tests on the Proportional Oculomotor Subscale Scores for VR Studies

	Study	113	725	651	650	106	103	105	104	112	110	109	108	101	102	111	107
Study	Mean	0.732	0.550	0.513	0.509	0.483	0.451	0.444	0.410	0.409	0.392	0.357	0.356	0.335	0.322	0.302	0.293
113	0.7322															*	*
725	0.5496																
651	0.5125																*
650	0.5091																
106	0.4829																
103	0.4508																
105	0.4439																*
104	0.4099																
112	0.4093																
110	0.3923																*
109	0.3568																
108	0.3564																
101	0.3352																
102	0.3218																
111	0.3020																
107	0.2932																

* Significant difference ($p < .05$)

The results of the post hoc analysis shown in Table 17 above indicated a significant difference in the mean scores of four HMD study comparisons and between one HMD and BOOM study. The direction of the difference in the proportional Oculomotor mean scores for each of the significant HMD study comparisons (cf. Table 14 or 17) shows: Study 113 (0.732) was greater than Study 111 (0.302); Study 113 (0.732) was also greater than Study 107 (0.293); and the means for Study 105 (0.444) and 110 (0.392) were both greater than Study 107 (0.293). For the HMD and BOOM study comparison, the mean of Study 651 (BOOM) was significantly greater than Study 107 (HMD), 0.513 and 0.293 respectively. Additionally, a review of Table 17 reveals that there were no significant differences on the proportional Oculomotor subscale for the comparisons between HMD and CAVE studies nor for the BOOM and CAVE study comparisons.

Table 18 provides a list of the equipment features that differed between each pair of studies identified as statistically different on this subscale, ordered from largest to smallest mean difference, as well as the associated significance level for each comparison. Again, caution is required when interpreting these results. Only five comparisons showed significantly different mean scores and, based on the error rate, six out of the 120 comparisons could have occurred merely by chance.

Table 18. Equipment Differences and Significance Levels for VR Studies with Significantly Different Mean Proportional Oculomotor Scores

Study Pair	VR Type	Mean Difference	Significance	Equipment Differences[1]
107, 113	HMD, HMD	0.4390	$p = .031$	TS
111, 113	HMD, HMD	0.4302	$p = .047$	TL, TS, W
107, 651	HMD, BOOM	0.2193	$p = .007$	D, F, R, T
105, 107	HMD, HMD	0.1507	$p = .004$	F, I, R, TS
107, 110	HMD, HMD	0.0990	$p = .010$	---

[1] D = Display Type, F = Field of View, I = IPD Adjust, R = Resolution, T = Head Tracking (Yes or No)
TS = Head Tracker Speed, W = Display Weight

The equipment features for the significant study comparisons shown in Table 18 indicate that the differences related to the head tracker were present in four of the study pairs. In particular, differences in the speed of the head tracker appear in three of the comparisons while head-tracking versus no head-tracking accounted for the other significant study pair. Display field-of-view and resolution also differed in two of the study comparisons (Studies 107 and 651 and Studies 105 and 107). In contrast, the type of display, IPD adjustability, and display weight each appeared as differences in equipment features in only one study comparison. A notable exception to the equipment differences between studies that had significantly different mean proportional Oculomotor scores was the comparison between Studies 107 and 110 which employed the same HMD. The implications of this result will be discussed in Chapter 5.

Finally, Table 19 provides a results summary for the post hoc analysis on the proportional Disorientation subscale scores for the VR studies.

Table 19. Results of the Multiple Comparison Tests on the Proportional Disorientation Subscale Scores for VR Studies

	Study	111	107	112	108	109	725	102	103	101	110	105	651	106	650	104	113
Study	Mean	0.454	0.410	0.378	0.350	0.348	0.346	0.342	0.341	0.334	0.330	0.311	0.279	0.276	0.263	0.259	0.106
111	0.4540																*
107	0.4098																*
112	0.3777																*
108	0.3503																*
109	0.3484																*
725	0.3455																
102	0.3421																
103	0.3408																
101	0.3339																
110	0.3299																*
105	0.3106																
651	0.2787																
106	0.2760																
650	0.2634																
104	0.2589																
113	0.1063																

* Significant difference ($p < .05$)

A review of Table 19 reveals a significant difference in means only for several of the HMD study comparisons. The results of the post hoc analysis indicated that Study 113 appeared in all of the significant study pairs. Moreover, the mean proportional Disorientation score for Study 113 (0.106) was significantly lower than the mean for every study to which it was compared. The results in Table 19 also indicate that the mean proportional Disorientation scores were not significantly different for the HMD and BOOM study comparisons as well as the BOOM and CAVE study comparisons.

Equipment feature differences between each pair of significantly different studies on the proportional Disorientation subscale and the associated significance level for each comparison is presented below in Table 20. As with the proportional Nausea and Oculomotor analyses, caution is required when interpreting these results due to the relatively small number of significant comparisons.

Table 20. Equipment Differences and Significance Levels for VR Studies with Significantly Different Mean Proportional Disorientation Scores

Study Pair	VR Type	Mean Difference	Significance	Equipment Differences[1]
111, 113	HMD, HMD	0.3478	$p = .004$	TL, TS, W
107, 113	HMD, HMD	0.3035	$p = .002$	TS
112, 113	HMD, HMD	0.2715	$p = .036$	TL, TS, W
108, 113	HMD, HMD	0.2440	$p = .012$	TS
109, 113	HMD, HMD	0.2421	$p = .013$	TS
110, 113	HMD, HMD	0.2236	$p = .024$	TS

[1] TL = Head Tracker Latency, TS = Head Tracker Speed, W = Display Weight

In all of the significant comparisons (cf. Table 20), differences in speed of the head tracker were noted. Additionally, two of these study pairs also differed in the latency of the head tracker and the weight of the display.

Analysis of Differences in Sickness Severity Between and Within System Types

Two additional sets of analyses were performed on the actual SSQ data (i.e., not the proportional variables) in order to evaluate differences in sickness severity both within and between VE system types as well as among the individual studies. The first set of analyses evaluated the SSQ Total Severity score to determine whether there were statistically significant differences in overall sickness severity whereas the second set of analyses evaluated scores on the SSQ subscales (Nausea, Oculomotor, and Disorientation) for simulator and VR systems.

Total Severity Score

An Independent t-test was used to assess whether there were differences in the Total Severity scores between simulators and VR devices. Table 21 shows that the overall level of sickness from exposure to simulators was significantly different than the sickness associated with exposure to VR devices ($p < .001$). Inspection of the two group means indicates that the average Total Severity score for VR devices was significantly greater than the average score for simulators.

Table 21. Comparison of Simulators and VR Devices on the SSQ Total Severity Score

Variable	n	*M*	*SD*	*t*	*df*	*p*
Total Severity Score						
Simulators	908	18.13	17.79	-10.09[a]	2062.8[a]	.000
VR Devices	1192	27.95	26.73			

[a] The *t* and *df* were adjusted because variances were not equal

A series of One-Way ANOVAs were then run on the Total Severity (TS) score. First, an ANOVA was conducted to determine whether there were differences in sickness severity between different types of simulators. The results revealed a statistically significant difference among the three types of simulator systems, $F(2, 905) = 20.78$, $p < .001$. Table 22 shows the means on the Total Severity score was 12.37 for Fixed-Wing, 17.12 for Rotary-Wing, and 23.34 for Driving simulators. Post hoc Games-Howell Tests indicated that there were significant mean differences between all of the simulator types. The mean Total Severity score for Fixed-Wing simulators was significantly lower than both Rotary-Wing ($p < .001$) and Driving ($p < .001$) simulators and the mean for Rotary-Wing was significantly less than Driving simulators ($p < .001$).

Table 22. Means and Standard Deviations of Total Severity Score for Three Types of Simulators

Simulator Type	n	M	SD
Fixed-Wing	150	12.37	11.10
Rotary-Wing	496	17.12	15.77
Driving	262	23.34	22.57
Total	908	18.13	17.79

Another ANOVA was then conducted to determine whether there were differences in overall sickness severity among the different simulator studies. The results revealed a statistically significant difference among the 21 simulator studies, $F(20, 887) = 10.25$, $p < .001$.

In Table 23, the means and standard deviations, grouped by type of simulator, are shown for each of the simulator studies.

Table 23. Means and Standard Deviations of Total Severity Score for Simulator Studies

Simulator Type	Study Number	n	M	SD
Driving	201	62	24.85	17.83
	202	53	43.26	30.89
	203	43	17.13	17.62
	204	104	14.85	14.14
Fixed-Wing	302	28	12.02	10.65
	303	18	23.89	16.58
	304	8	6.55	5.19
	306	10	8.98	5.05
	307	39	12.27	10.72
	308	20	11.97	8.28
	316	19	9.45	8.02
	318	8	6.08	3.97
Rotary-Wing	305	86	14.35	13.23
	309	66	12.24	11.56
	310	67	20.60	19.19
	311	42	16.12	11.61
	312	125	22.29	18.51
	313	38	17.03	16.19
	314	30	15.21	11.70
	315	28	11.22	11.38
	317	14	13.36	13.34
	Total	908	18.13	17.79

Since the results of the ANOVA were significant, post hoc Games-Howell Tests were used to identify which study pairs had significant differences in their mean Total Severity scores. Table 24 below provides a summary of the multiple comparison test results. The data within the table are arrayed in the same manner as the post hoc tests results shown previously for the proportional subscales scores. That is, the data is ordered according to the magnitude of the mean score, from highest to lowest, along the horizontal and vertical axes and a significant

difference between two studies is indicated with an asterisk (*) in the cell corresponding to the intersection of the two studies, whereas non-significant study pairs are empty.

Table 24. Results of the Multiple Comparison Tests on the Total Severity Score for Simulator Studies

	Study	202	201	303	312	310	203	313	311	314	204	305	317	307	309	302	308	315	316	306	304	318
Study	Mean	43.3	24.9	23.9	22.3	20.6	17.1	17.0	16.1	15.2	14.9	14.4	13.4	12.3	12.2	12.0	12.0	11.2	9.5	9.0	6.6	6.1
202	43.3		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
201	24.9										*	*		*	*	*	*	*	*	*	*	*
303	23.9																			*	*	*
312	22.3										*	*		*	*	*	*	*	*	*	*	*
310	20.6																		*	*	*	*
203	17.1																					*
313	17.0																					*
311	16.1																				*	*
314	15.2																					*
204	14.9																				*	*
305	14.4																					*
317	13.4																					
307	12.3																					
309	12.2																					
302	12.0																					
308	12.0																					
315	11.2																					
316	9.5																					
306	9.0																					
304	6.6																					
318	6.1																					

The results in Table 24 reveal that Driving simulator Study 202 was significantly different than all of the other studies. A comparison of the means indicates that this particular study not only had the highest mean Total Severity score, it was almost twice as high as the next largest mean in Study 201, 43.26 and 24.85 respectively. In addition to the mean score for the Driving simulator in Study 202 having a significantly higher mean than the scores for the other Driving simulator studies (i.e., Studies 202, 203, and 204), Table 24 also shows that the mean for the Driving simulator in Study 201 (24.85) was also significantly greater than the Driving simulator in Study 204 (14.85). The results of the post hoc analyses for the other within simulator type comparisons indicate that the mean Total Severity Score for the Fixed-Wing simulator in Study 303 (23.89) was significantly greater than the Fixed-Wing simulators in Studies 304 (6.55), 306 (8.98) and 318 (6.08) and the Rotary-Wing simulator in Study 312 (mean = 22.29) was significantly greater than the Rotary-Wing simulators in Studies 305 (14.35), 309 (12.24) and 315 (11.22).

The post hoc tests also revealed mean differences in Total Severity scores for the between simulator type comparisons. An obvious result, based on a visual inspection of the mean scores, was that the Driving simulator in Study 202 showed a significantly greater mean than all of the Rotary- and Fixed-Wing simulators (c.f. Table 24 above). The other significant Driving and Rotary-Wing comparisons reveal that the mean Total Severity score for the Driving simulator in Study 201 (24.85) was significantly greater than the mean scores for Rotary-Wing Studies 305 (14.35), 309 (12.24) and 315 (11.22). Conversely, the Driving simulator in Study 204 (14.85) showed a significantly lower mean than the Rotary-Wing simulator in Study 312 (22.29). In the significant Driving and Fixed-Wing comparisons, the results in Table 24 show that the mean score for Driving simulator Study 201 (24.85) was significantly greater than all but one of the

Fixed-Wing studies. The exception was Study 303 that had a mean score of 23.89. Additionally, the Driving simulator in Study 204 had a significantly larger mean than the Fixed-Wing simulators in Studies 304 and 318, the means of which reflect a relatively negligible overall sickness score (6.55 and 6.08, respectively). The mean for Driving simulator Study 203 (17.13) was also significantly greater than Study 318 (6.08).

Finally, the results for the Rotary- and Fixed-Wing comparisons indicate that the Rotary-Wing simulator in Study 312 had a significantly greater mean (22.29) than all of the Fixed-Wing simulators except Study 303 that had a mean Total Severity score of 22.89 which was not significantly different (c.f. Table 23 above). The mean for Rotary-Wing Study 310 (20.60) was also significantly greater than the Fixed-Wing simulators in Studies 304 (6.55), 306 (8.98), 316 (9.45), and 318 (6.08). Relatedly, Fixed-Wing Study 318 had a significantly lower mean (6.08) than the Rotary-Wing simulators in Studies 305 (14.35), 311 (16.12), 313 (17.03), and 314 (15.21) and the mean for Study 304 (6.55; Fixed-Wing) was also significantly less than Study 311 (16.12; Rotary-Wing). Additional information on the mean differences in Total Severity score and significance level for each of the significant simulator study comparisons is provided in Appendix F.

Next, a One-Way ANOVA was conducted to determine whether there were differences in total sickness severity between different types of VR devices. The results revealed a statistically significant difference among the three types of VR devices, $F(2, 1189) = 13.99$, $p < .001$. The means and standard deviations for the different VR devices are shown below in Table 25. Post hoc Games-Howell Tests indicated that there were significant mean differences between all of the VR devices. The mean Total Severity score for HMDs was significantly greater BOOM ($p =$

.046) and CAVE ($p < .001$) systems and the mean for BOOMs was significantly greater than CAVEs ($p < .001$).

Table 25. Means and Standard Deviations of Total Severity Score for Three Types of VR Devices

VR System	n	M	SD
HMD	1100	28.97	27.01
BOOM	57	21.32	22.89
CAVE	35	6.63	4.26
Total	1192	27.95	26.73

Lastly, a One-Way ANOVA was conducted to determine whether there were differences in total sickness severity among the 16 VR studies. The results revealed a statistically significant difference among the various VR studies, $F(15, 1176) = 7.82$, $p < .001$. Table 26 shows the descriptive statistics for each of the VR studies, grouped by type of VR device.

Table 26. Means and Standard Deviations of Total Severity Score for VR Studies

VR System	Study Number	n	M	SD
HMD	101	47	19.26	14.94
	102	13	52.36	32.14
	103	25	23.64	25.66
	104	19	24.8	34.55
	105	81	36.06	34.98
	106	30	29.55	27.66
	107	200	25.3	26.21
	108	197	29.71	25.82
	109	194	32.95	26.96
	110	211	34.26	26.71
	111	32	11.45	12.81
	112	39	10.26	7.56
	113	12	9.66	7.56
BOOM	650	25	26.63	30.92
	651	32	17.18	12.88
CAVE	725	35	6.63	4.26
	Total	1192	27.95	26.73

Post hoc Games-Howell Tests were again used to identify which study pairs had significant differences in their mean Total Severity scores. Table 27 below provides a summary of the multiple comparison test results.

Table 27. Results of the Multiple Comparison Tests on the Total Severity Score for VR Studies

	Study	102	105	110	109	108	106	650	107	104	103	101	651	111	112	113	725
Study	Mean	52.36	36.06	34.26	32.95	29.71	29.55	26.63	25.30	24.80	23.64	19.26	17.18	11.45	10.26	9.66	6.63
102	52.36													*	*	*	*
105	36.06											*	*	*	*	*	*
110	34.26											*	*	*	*	*	*
109	32.95											*	*	*	*	*	*
108	29.71											*	*	*	*	*	*
106	29.55														*		*
650	26.63																
107	25.30													*	*	*	*
104	24.80																
103	23.64																
101	19.26														*		*
651	17.18																*
111	11.45																
112	10.26																
113	9.66																
725	6.63																

The results in Table 27 reveal that the only significant difference within VR system type comparisons were with the HMDs. Since data for CAVE-type systems were only available for one study, obviously no within system comparisons were possible, but the results for the two BOOM studies indicated that their mean Total Severity scores (26.63 and 17.18) were not significantly different. In contrast, 31% of the within system HMD comparisons revealed a significant difference in mean Total Severity. The results for these comparisons showed that the mean for the HMD systems in Studies 105 (36.06), 108 (29.71), 109 (32.95), and 110 (34.26) were each significantly greater than the mean scores for the HMDs in Studies 101 (19.26), 111 (11.45), 112 (10.26), and 113 (9.66). Likewise, the mean Total Severity score for the HMDs in Studies 102 (52.36) and 107 (25.30) were both significantly greater than Studies 111, 112, and 113, but the mean score in Studies 101 (19.26) and 106 (29.55) were only significantly greater than Study 112.

The results for the between VR system comparisons indicated that the mean Total Severity score for the BOOM system in Study 651 (17.18) was significantly lower than four of the HMD studies (Studies 105, 108, 109, and 110), but significantly greater than the mean for the CAVE study (6.63). Conversely, the mean for the BOOM system in Study 650 (26.63) was not significantly different than any of the HMD studies nor the CAVE study. Finally, the comparisons between the HMD and CAVE systems revealed that the CAVE system had a significantly lower mean (6.63) than 62% (8 out of 13) of the HMD studies.

SSQ Subscale Scores for Simulators

In order to determine whether there were differences in sickness severity on the individual SSQ subscales between various types of Simulator system configurations, three One-

Way ANOVAs were run on the subscale scores (one each for Nausea, Oculomotor, Disorientation) from 21 simulator studies, where each study represented a homogenous set of equipment features. The results of the analysis on the Nausea subscale revealed a statistically significant difference among the discrete simulator studies, $F(20, 887) = 5.54$, $p < .001$. Table 28 presents the descriptive statistics (i.e., means and standard deviations) of the Nausea subscale scores for each of the simulator studies, which are grouped by type of simulator system.

Table 28. Means and Standard Deviations of the SSQ Nausea Subscale Score for Studies Using Three Types of Simulator Systems

			Nausea Score	
Simulator Type	Study Number	n	M	SD
Driving	201	62	22.62	21.31
	202	53	31.5	28.60
	203	43	12.65	17.75
	204	104	13.39	17.00
Fixed-Wing	302	28	12.95	12.24
	303	18	16.43	15.97
	304	8	5.96	7.10
	306	10	6.68	7.85
	307	39	4.89	7.22
	308	20	10.97	11.69
	316	19	8.03	9.68
	318	8	2.38	4.42
Rotary-Wing	305	86	9.76	14.56
	309	66	12	14.62
	310	67	15.66	19.13
	311	42	12.27	11.97
	312	125	18.85	20.40
	313	38	14.81	19.30
	314	30	10.81	13.43
	315	28	8.86	10.99
	317	14	9.54	16.31
	Total	908	14.52	18.24

Since the results of the ANOVA were significant, post hoc Games-Howell Tests were used to identify which study pairs had significant differences in their mean Nausea scores. Table 29 below provides a summary of the multiple comparison test results.

Table 29. Results of the Multiple Comparison Tests on the Nausea Subscale Scores for Simulator Studies

	Study	202	201	312	303	310	313	204	302	203	311	309	308	314	305	317	315	316	306	304	307	318
Study	Mean	31.5	22.6	18.9	16.4	15.7	14.8	13.4	13.0	12.7	12.3	12.0	11.0	10.8	9.8	9.5	8.9	8.0	6.7	6.0	4.9	2.4
202	31.50					*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
201	22.62											*			*		*	*	*	*	*	*
312	18.85														*		*	*	*	*	*	*
303	16.43																					
310	15.66																				*	*
313	14.81																					*
204	13.39																				*	*
302	12.95																					*
203	12.65																					
311	12.27																				*	*
309	12.00																				*	*
308	10.97																					
314	10.81																					
305	9.76																					
317	9.54																					
315	8.86																					
316	8.03																					
306	6.68																					
304	5.96																					
307	4.89																					
318	2.38																					

Table 30 provides a list of the equipment features that differed between each pair of studies shown above in Table 29, which were identified as statistically different on the Nausea subscale, and the associated significance level for each comparison. The significant study pairs in the table are ordered from largest to smallest mean difference.

Table 30. Equipment Differences and Significance Levels for Simulator Studies with Significantly Different Mean Nausea Scores

Study Pair	Simulator Type	Mean Difference	Significance	Equipment Differences[1]
202, 203	Driving, Driving	18.854	$p = .022$	D, F, L, M, R
202, 204	Driving, Driving	18.107	$p = .010$	D, F, M, R
202, 315	Driving, Rotary	22.641	$p < .001$	F, L, M, S
202, 317	Driving, Rotary	21.960	$p = .063$	F, M, S
202, 305	Driving, Rotary	21.738	$p < .001$	F, L, M, S
202, 314	Driving, Rotary	20.688	$p = .004$	F, M, S
202, 309	Driving, Rotary	19.503	$p = .004$	F, M, S
202, 311	Driving, Rotary	19.234	$p = .005$	F, L, M, S
202, 313	Driving, Rotary	16.688	$p = .131$	F, M, S
202, 310	Driving, Rotary	15.837	$p = .091$	F, L, M, S
201, 315	Driving, Rotary	13.760	$p = .017$	D, F, L, M, S
201, 305	Driving, Rotary	12.857	$p = .012$	D, F, L, M, S
201, 309	Driving, Rotary	10.622	$p = .146$	D, F, M, S
202, 318	Driving, Fixed	29.115	$p < .001$	D, F, M, S
202, 307	Driving, Fixed	26.608	$p < .001$	F, L, M, S
202, 304	Driving, Fixed	25.537	$p < .001$	D, F, S
202, 306	Driving, Fixed	24.822	$p < .001$	F, S
202, 316	Driving, Fixed	23.466	$p < .001$	D, F, M, S
202, 308	Driving, Fixed	20.529	$p = .007$	F, M, S
201, 318	Driving, Fixed	20.234	$p < .001$	D, F, M, S
202, 302	Driving, Fixed	18.553	$p = .016$	D, F, S
201, 307	Driving, Fixed	17.727	$p < .001$	D, F, L, M, S
201, 304	Driving, Fixed	16.657	$p = .012$	D, F, S
201, 306	Driving, Fixed	15.941	$p = .014$	D, F, S
201, 316	Driving, Fixed	14.585	$p = .013$	D, F, M, S
204, 318	Driving, Fixed	11.008	$p = .005$	D, F, MD, S
204, 307	Driving, Fixed	8.500	$p = .008$	D, F, MD, S
312, 318	Rotary, Fixed	16.466	$p < .001$	D, F, S

307, 312	Fixed, Rotary	13.959	$p < .001$	F, L, S
310, 318	Rotary, Fixed	13.278	$p = .003$	D, F, S
304, 312	Fixed, Rotary	12.889	$p = .054$	D, F, M, S
313, 318	Rotary, Fixed	12.427	$p = .092$	D, F, S
306, 312	Fixed, Rotary	12.173	$p = .061$	F, M, S
312, 316	Rotary, Fixed	10.817	$p = .052$	D, F, S
307, 310	Fixed, Rotary	10.770	$p = .012$	F, L, S
311, 318	Rotary, Fixed	9.881	$p = .032$	D, F, S
309, 318	Rotary, Fixed	9.612	$p = .035$	D, F, S
307, 311	Fixed, Rotary	7.373	$p = .119$	F, L, S
307, 309	Fixed, Rotary	7.105	$p = .128$	F, S
302, 318	Fixed, Fixed	10.562	$p = .063$	F, M
312, 315	Rotary, Rotary	9.992	$p = .064$	F, L
305, 312	Rotary, Rotary	9.089	$p = .030$	F, L

[1] D = Display Type, F = Field of View, L = System Latency, M = Motion Base, MD = Motion Base DOF, R = Resolution (driving sims only), S = Simulator Type

The equipment features for the significant study comparisons shown in Table 30 indicate that overall, differences related to the display's field-of-view were present in all of the study pairs. Additionally, the type of display and motion base accounted for equipment differences in 55% of the significant comparisons between studies. On the other hand, overall system latency was a noted equipment difference in 31% of the significant study comparisons.

In terms of the individual study comparisons, the results in Table 29 indicate that Driving simulator Study 202 was significantly different than two of the other Driving simulator studies. A comparison of the means for these studies reveal that the mean Nausea score for Study 202 (31.50) was greater than the mean for Study 203 (12.65) and Study 204 (13.39). As shown in Table 30, the equipment differences in both of these study pairs included the type of display, resolution of the display, field-of-view, and whether the simulator provided motion (i.e., the motion base). The results of the other significant within system type comparisons revealed that only one of the Fixed-Wing study pairs differed; the mean for Study 302 (12.95) was

significantly greater than Study 318 (2.38) and there were only two significant Rotary-Wing study pairs. In both of these comparisons, Study 312 had a significantly greater mean (18.85) than Study 305 (9.76) and Study 315 (8.86). Table 30 shows that the significant Fixed-Wing comparison differed in terms of the field-of-view and the motion base, whereas both of the Rotary-Wing pairs had a different field-of-view and a different overall system latency.

The results for the between system type comparisons shown in Table 29 revealed that the mean Nausea score for the Driving simulator in Study 202 (31.50) was significantly greater than that for all of the Rotary-Wing simulators except Study 312 (18.85). Table 30 indicates that the equipment for Study 202 differed from every significant Rotary-Wing comparison on the field-of-view, motion base, and obviously, on the system type. The overall system latency also differed in half of these significant comparisons. Driving simulator in Study 201 also showed a significantly greater mean (22.62) than three of the Rotary-Wing studies: Study 305 (mean = 9.76), Study 309 (mean = 12.00), and Study 315 (mean = 8.860. In addition to the expected difference in system type, the equipment for Study 201 differed from all three significant Rotary-Wing comparisons on the display type, field-of-view, and motion base. Overall system latency was also a noted equipment difference in two of the three significant comparisons.

Significant differences were also found in the between system type comparisons for the Driving and Fixed-Wing study pairs. The mean Nausea score for the Driving simulator Study 202 was significantly greater than all of the Fixed-Wing simulators except Study 303 (16.43). Similar to this study's equipment comparison with the Rotary-Wing simulators, the field-of-view (and of course the system type) differed from all of the significant Fixed-Wing simulators. Differences in the display type and motion base also appeared as a factor in half of the significant comparisons, whereas the overall system latency was only different in one of the

study pairs (Studies 202 and 307). Moreover, the Driving simulator in Study 201 had a significantly greater Nausea score mean (22.62) than a little more than half (63%) of the Fixed-Wing simulators. In each of these significant study pairs, the simulators had different types of displays and fields-of-view (cf. Table 30). Differences in simulated motion were noted in three of the five significant comparisons, but overall system latency only differed in one of the study pairs (Study 201 and 307). The results in Table 29 also indicated that the mean for Driving simulator Study 204 was significantly greater than Fixed-Wing Study 307 (4.89) and Study 318 (2.38), which had relatively negligible Nausea symptoms. The equipment differences for both of these significant study pairs were the type of display, field-of-view, and the degrees of freedom (direction of simulated motion) provided by the motion base (cf. Table 30).

The results for the Fixed- and Rotary-Wing study comparisons indicated that the Rotary-Wing simulator used in Study 312 had the highest mean Nausea score out of all the flight simulators (18.85), which was significantly greater than 63% of the Fixed-Wing simulators (Studies 304, 306, 307, 316, and 318). Differences in field-of-view were noted for all of the significant comparisons between Study 312 and the Fixed-Wing simulators (cf. Table 30). Additionally, differences in the display type occurred in 60% of these comparisons and simulated motion differed in two of the five comparisons. However, a difference in overall system latency was only noted in one of the study pairs (Study 307 and 312).

In contrast to the mean for the Rotary-Wing Study 312, the Fixed-Wing simulator in Study 318 had the lowest mean (2.38) for all of the flight simulators and thus, was significantly lower than about half (56%) of the Rotary-Wing studies. These significant study comparisons all differed on the type of display and the field-of-view. Not surprisingly, the results for this Fixed-Wing simulator were not significantly different than the Rotary-Wing simulators that also had

relatively low mean Nausea scores (i.e., Studies 305, 314, 315, and 317 which all had a mean score less than 11).

The next ANOVA was conducted on the Oculomotor subscale scores for all of the simulator studies. The results of the analysis revealed a statistically significant difference among the 21 simulator studies, $F(20, 887) = 8.55$, $p < .001$. In Table 31, the Oculomotor subscale score means and standard deviations, grouped by type of simulator, are shown for each of the simulator studies.

Table 31. Means and Standard Deviations of the SSQ Oculomotor Subscale Score for Studies Using Three Types of Simulator Systems

			Oculomotor Score	
Simulator Type	Study Number	n	M	SD
Driving	201	62	17.61	12.56
	202	53	36.76	27.23
	203	43	13.22	12.61
	204	104	12.39	10.13
Fixed-Wing	302	28	9.75	10.89
	303	18	26.11	14.29
	304	8	6.63	2.68
	306	10	11.37	6.44
	307	39	18.27	13.44
	308	20	13.64	9.71
	316	19	10.37	11.64
	318	8	8.53	6.33
Rotary-Wing	305	86	16.39	13.30
	309	66	11.71	10.82
	310	67	20.82	16.97
	311	42	16.96	12.51
	312	125	21.65	17.02
	313	38	16.96	13.12
	314	30	17.94	12.82
	315	28	12.45	10.96
	317	14	15.16	11.12
	Total	908	17.4	15.55

Post hoc Games-Howell Tests were then used to identify which study pairs had significant differences in their mean Oculomotor scores. Table 32 below provides a summary of the multiple comparison test results.

Table 32. Results of the Multiple Comparison Tests on the Oculomotor Subscale Scores for Simulator Studies

	Study	202	303	312	310	307	314	201	311	313	305	317	308	203	315	204	309	306	316	302	318	304
Study	Mean	36.8	26.1	21.7	20.8	18.3	17.9	17.6	17.0	17.0	16.4	15.2	13.6	13.2	12.5	12.4	11.7	11.4	10.4	9.8	8.5	6.6
202	36.76			*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
303	26.11														*	*	*	*	*	*	*	*
312	21.65													*	*	*	*	*	*	*	*	*
310	20.82															*	*			*	*	*
307	18.27																					*
314	17.94																					*
201	17.61																					*
311	16.96																					*
313	16.96																					*
305	16.39																					*
317	15.16																					
308	13.64																					
203	13.22																					
315	12.45																					
204	12.39																					*
309	11.71																					
306	11.37																					
316	10.37																					
302	9.75																					
318	8.53																					
304	6.63																					

Table 33 provides a list of the equipment features that differed between each pair of studies shown above in Table 32, which were identified as statistically different on the Oculomotor subscale, and the associated significance level for each comparison. The significant study pairs in the table are ordered from largest to smallest mean difference.

Table 33. Equipment Differences and Significance Levels for Simulator Studies with Significantly Different Mean Oculomotor Scores

Study Pair	Simulator Type	Mean Difference	Significance	Equipment Differences[1]
202, 204	Driving, Driving	24.365	$p < .001$	D, F, M, R
202, 203	Driving, Driving	23.535	$p < .001$	D, F, L, M, R
201, 202	Driving, Driving	19.151	$p = .002$	D, F, L, R
202, 309	Driving, Rotary	25.041	$p < .001$	F, M, S
202, 315	Driving, Rotary	24.303	$p < .001$	F, L, M, S
202, 317	Driving, Rotary	21.596	$p = .005$	F, M, S
202, 305	Driving, Rotary	20.362	$p = .001$	F, L, M, S
202, 313	Driving, Rotary	19.801	$p = .003$	F, M, S
202, 311	Driving, Rotary	19.791	$p = .002$	F, L, M, S
202, 314	Driving, Rotary	18.817	$p = .008$	F, M, S
202, 310	Driving, Rotary	15.939	$p = .045$	F, L, M, S
202, 312	Driving, Rotary	15.107	$p = .046$	F, L, M, S
204, 312	Driving, Rotary	9.258	$p < .001$	D, F, MD, S
203, 312	Driving, Rotary	8.428	$p = .096$	D, F, L, S
204, 310	Driving, Rotary	8.426	$p = .051$	D, F, MD, S
202, 304	Driving, Fixed	30.123	$p < .001$	D, F, S
202, 318	Driving, Fixed	28.228	$p < .001$	D, F, M, S
202, 302	Driving, Fixed	27.010	$p < .001$	D, F, S
202, 316	Driving, Fixed	26.383	$p < .001$	D, F, M, S
202, 306	Driving, Fixed	25.386	$p < .001$	F, S
202, 308	Driving, Fixed	23.112	$p < .001$	F, M, S
202, 307	Driving, Fixed	18.486	$p = .008$	F, L, M, S
204, 303	Driving, Fixed	13.719	$p = .069$	D, F, MD, S
201, 304	Driving, Fixed	10.973	$p < .001$	D, F, S
204, 304	Driving, Fixed	5.758	$p = .026$	D, F, M, S
304, 312	Fixed, Rotary	15.016	$p < .001$	D, F, M, S
303, 309	Fixed, Rotary	14.394	$p = .054$	F, S
304, 310	Fixed, Rotary	14.184	$p < .001$	D, F, M, S

303, 315	Fixed, Rotary	13.656	$p = .134$	F, S
312, 318	Rotary, Fixed	13.121	$p = .018$	D, F, S
310, 318	Rotary, Fixed	12.289	$p = .049$	D, F, S
302, 312	Fixed, Rotary	11.903	$p = .003$	D, F, M, S
304, 314	Fixed, Rotary	11.307	$p = .010$	D, F, M, S
312, 316	Rotary, Fixed	11.276	$p = .083$	D, F, S
302, 310	Fixed, Rotary	11.071	$p = .038$	D, F, M, S
304, 311	Fixed, Rotary	10.332	$p = .003$	D, F, M, S
304, 313	Fixed, Rotary	10.323	$p = .009$	D, F, M, S
306, 312	Fixed, Rotary	10.278	$p = .050$	F, M, S
304, 305	Fixed, Rotary	9.761	$p < .001$	D, F, M, S
303, 304	Fixed, Fixed	19.476	$p = .002$	D, F, M
303, 318	Fixed, Fixed	17.581	$p = .023$	D, F
302, 303	Fixed, Fixed	16.363	$p = .029$	D, F, M
303, 316	Fixed, Fixed	15.736	$p = .082$	D, F
303, 306	Fixed, Fixed	14.739	$p = .080$	F, M
304, 307	Fixed, Fixed	11.637	$p = .002$	D, F, M
309, 312	Rotary, Rotary	9.934	$p < .001$	F
312, 315	Rotary, Rotary	9.196	$p = .076$	F, L
309, 310	Rotary, Rotary	9.102	$p = .045$	- - -

[1] D = Display Type, F = Field of View, L = System Latency, M = Motion Base, MD = Motion Base DOF, R = Resolution (driving sims only), S = Simulator Type

The equipment features for the significant study comparisons shown in Table 33 indicate that overall, display field-of-view again differed for all of the significant comparisons on the Oculomotor subscale with the exception of Study pair 309 and 310. A comparison of the equipment features for these two Rotary-Wing simulators indicated that there were no differences between them. The implication of this finding will be addressed in Chapter 5. In addition to differences in field-of-view, the type of display and motion base (i.e., motion versus no motion) were present in 60% of the significant comparisons. Although, differences in overall system latency was only noted in 21% of the significant comparisons.

In terms of the individual study comparisons, the results in Table 32 indicate that the mean Oculomotor score for Driving simulator Study 202 (36.76) was significantly greater than

all of the other Driving simulators. Moreover, the mean for this study was more than double the mean of the next highest score for all of the Driving simulators. As shown in Table 33, the equipment differences in all three of the significant Driving simulator comparisons included the type of display, field-of-view, and display resolution. Differences in whether the simulator provided motion (i.e., the motion base) were also noted in two of the study pairs (Study 202 and 203 and Study 202 and 204), whereas the overall system latency differed in the other pair (Study 201 and 202).

The results of the other significant within system type comparisons revealed that the mean Oculomotor score for the Fixed-Wing simulator in Study 303 (26.11) was significantly greater than most of the other Fixed-Wing studies. The two exceptions were Study 307 (18.27) and Study 308 (13.64), which were not significantly different. Study 307 also had a significantly greater mean than Study 304 (18.27 and 6.63, respectively). Table 33 shows that all of the significant Fixed-Wing comparison differed in terms of their field-of-view. Differences in the type of display were also noted in four of the significant comparisons with Study 303 as well as the comparison between Study 304 and 307. Furthermore, the motion base differed in three of the significant comparisons with Study 303 and in the Study 304 and 307 pair.

In the Rotary-Wing study comparisons, the analysis only revealed three significant pairs. In two of these comparisons, Study 312 had a significantly greater mean (21.65) than Study 309 (11.71) and Study 315 (12.45). The other significant Rotary-Wing pair showed that the mean for Study 310 (20.82) was also significantly greater than Study 309. Both of the comparisons with Study 312 had differences in the field-of-view while Study pair 312 and 315 also differed in overall system latency. However, there were no differences in equipment features noted for Study 309 and 310, which will be discussed in Chapter 5.

The results for the between system type comparisons shown in Table 32 revealed that the mean Oculomotor score for the Driving simulator in Study 202 (36.76) was significantly greater than all of the Rotary-Wing simulators. Table 33 indicates that the equipment for Study 202 differed from every significant Rotary-Wing comparison on the field-of-view and motion base, whereas differences in system latency occurred in 56% of the comparisons. In contrast, the mean of the Rotary-Wing simulator in Study 312 (21.65) was significantly greater than the Driving simulators in Study 203 (13.22) and Study 204 (12.39), but not significantly different than the Driving simulator in Study 201 (17.61). Differences in equipment for both of the significant study pairs included field-of-view and display type (cf. Table 33). Differences in system latency were also noted for the comparison between Study 203 and 312, whereas differences in the degrees of freedom of the motion base occurred in the Study 204 and 312 comparison. Finally, the Rotary-Wing simulator in Study 310 also had a significantly greater mean (20.82) than the Driving simulator in Study 204 (12.39). The equipment in these two studies also differed in terms of the type of display, field-of-view, and degrees of freedom in the motion base.

Significant differences were also found in the between system type comparisons for the Driving and Fixed-Wing study pairs. The mean Oculomotor score for the Driving simulator Study 202 (36.76) was significantly greater than all of the Fixed-Wing simulators (except Study 303, which was not significantly different) and a noted equipment difference in all of these significant comparisons was the field-of-view. Differences in the display type and motion base also appeared as a factor in more than half (57%) of the significant comparisons, but overall system latency was only noted in one of the six significant comparisons. Also, the means for the Driving simulators in Study 201 (17.61) and Study 204 (12.39) were significantly greater than

the Fixed-Wing simulator in Study 304 (6.63). Conversely, Study 303 was the only Fixed-Wing simulator that showed a significantly higher mean than any of the Driving simulators, which appeared in the comparison with Study 204 (26.11 and 13.39, respectively). The equipment differences for all three of these significant study pairs were the type of display and field-of-view. Additionally, the comparisons between Study 204 and 304 differed in whether simulated motion was present, whereas the Study 303 and 204 differed in the degrees of freedom of the motion base (cf. Table 33).

The results for the Fixed- and Rotary-Wing study comparisons indicated that the Rotary-Wing simulator used in Study 312 had a significantly greater Oculomotor mean than 63% of the Fixed-Wing simulators (Studies 302, 304, 306, 316, and 318) and differences in field-of-view were noted for all of them (cf. Table 33). Other noted equipment differences included display type (80% of the comparisons) and motion base (60%). In addition to the comparison between Study 304 and 312, the means for five other Rotary-Wing studies were also significantly greater than the mean for Study 304, which had the smallest mean Oculomotor score out of all of the Fixed-Wing simulators (6.63). These significant comparisons all differed in the display type, field-of-view, and motion base. The mean for Rotary-Wing Study 310 (20.82) was also significantly greater than the Fixed-Wing in Study 318 (8.53) and had different types of displays and fields-of-view. In contrast to these findings, two of the Fixed-Rotary comparisons showed an opposite directional difference in Oculomotor score means. Specifically, the Fixed-Wing simulator in Study 303 had a significantly greater mean (26.11) than the Rotary-Wing simulators in Studies 309 (11.71) and 315 (12.75) and the field-of-view differed in both study pairs.

Finally, a One-Way ANOVA was run on the Disorientation subscale scores from the 21 simulator studies. The results of the analysis revealed a statistically significant difference among

the discrete simulator studies, $F(20, 887) = 12.78$, $p < .001$. The descriptive statistics (i.e., means and standard deviations) of the Disorientation subscale scores for each of the simulator studies, grouped by type of simulator system, are presented in Table 34.

Table 34. Means and Standard Deviations of the SSQ Disorientation Subscale Score for Studies Using Three Types of Simulator Systems

			Disorientation Score	
Simulator Type	Study Number	n	M	SD
Driving	201	62	27.17	25.26
	202	53	47.54	36.57
	203	43	21.04	24.93
	204	104	12.98	17.81
Fixed-Wing	302	28	7.95	12.81
	303	18	17.01	23.66
	304	8	3.48	9.84
	306	10	2.78	5.87
	307	39	5	10.34
	308	20	3.48	8.89
	316	19	4.4	8.11
	318	8	3.48	6.44
Rotary-Wing	305	86	9.06	13.80
	309	66	6.54	11.51
	310	67	15.58	22.93
	311	42	10.94	14.27
	312	125	15.7	20.46
	313	38	10.62	16.32
	314	30	7.89	10.13
	315	28	5.97	12.81
	317	14	7.95	10.52
	Total	908	14.33	21.57

Games-Howell Tests were then used for the post hoc multiple comparisons to identify which study pairs had significant differences in their mean Disorientation scores. Table 35 below provides a summary of the multiple comparison test results.

Table 35. Results of the Multiple Comparison Tests on the Disorientation Subscale Scores for Simulator Studies

	Study	202	201	203	303	312	310	204	311	313	305	302	317	314	309	315	307	316	304	308	318	306
Study	Mean	47.5	27.2	21.0	17.0	15.7	15.6	13.0	10.9	10.6	9.1	8.0	8.0	7.9	6.5	6.0	5.0	4.4	3.5	3.5	3.5	2.8
202	47.54		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
201	27.17							*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
203	21.04														*	*	*	*		*	*	*
303	17.01																					
312	15.70														*		*	*		*	*	*
310	15.58																	*		*		*
204	12.98																*	*		*		*
311	10.94																					
313	10.62																					
305	9.06																					
302	7.95																					
317	7.95																					
314	7.89																					
309	6.54																					
315	5.97																					
307	5.00																					
316	4.40																					
304	3.48																					
308	3.48																					
318	3.48																					
306	2.78																					

Table 36 provides a list of the equipment features that differed between each pair of studies that were identified in Table 35 as statistically different on the Disorientation subscale and the associated significance level for each comparison. The significant study pairs in the table are ordered from largest to smallest mean difference.

Table 36. Equipment Differences and Significance Levels for Simulator Studies with Significantly Different Mean Disorientation Scores

Study Pair	Simulator Type	Mean Difference	Significance	Equipment Differences[1]
202, 204	Driving, Driving	34.555	$p < .001$	D, F, M, R
202, 203	Driving, Driving	26.496	$p = .009$	D, F, L, M, R
201, 202	Driving, Driving	20.372	$p = .102$	D, F, L, R
201, 204	Driving, Driving	14.183	$p = .026$	F, M, R
202, 315	Driving, Rotary	41.572	$p < .001$	F, L, M, S
202, 309	Driving, Rotary	41.000	$p < .001$	F, M, S
202, 314	Driving, Rotary	39.650	$p < .001$	F, M, S
202, 317	Driving, Rotary	39.584	$p < .001$	F, M, S
202, 305	Driving, Rotary	38.474	$p < .001$	F, L, M, S
202, 313	Driving, Rotary	36.915	$p < .001$	F, M, S
202, 311	Driving, Rotary	36.601	$p < .001$	F, L, M, S
202, 310	Driving, Rotary	31.956	$p < .001$	F, L, M, S
202, 312	Driving, Rotary	31.836	$p < .001$	F, L, M, S
201, 315	Driving, Rotary	21.201	$p < .001$	D, F, L, M, S
201, 309	Driving, Rotary	20.628	$p < .001$	D, F, M, S
201, 314	Driving, Rotary	19.278	$p < .001$	D, F, M, S
201, 317	Driving, Rotary	19.212	$p = .006$	D, F, M, S
201, 305	Driving, Rotary	18.102	$p < .001$	D, F, L, M, S
201, 313	Driving, Rotary	16.543	$p = .019$	D, F, M, S
201, 311	Driving, Rotary	16.229	$p = .010$	D, F, L, M, S
203, 315	Driving, Rotary	15.076	$p = .132$	D, F, L, S
203, 309	Driving, Rotary	14.504	$p = .080$	D, F, S
202, 306	Driving, Fixed	44.754	$p < .001$	F, S
202, 304	Driving, Fixed	44.058	$p < .001$	D, F, S
202, 308	Driving, Fixed	44.058	$p < .001$	F, M, S
202, 318	Driving, Fixed	44.058	$p < .001$	D, F, M, S
202, 316	Driving, Fixed	43.142	$p < .001$	D, F, M, S
202, 307	Driving, Fixed	42.541	$p < .001$	F, L, M, S
202, 302	Driving, Fixed	39.584	$p < .001$	D, F, S

202, 303	Driving, Fixed	30.525	$p = .023$	F, M, S
201, 306	Driving, Fixed	24.382	$p < .001$	D, F, S
201, 304	Driving, Fixed	23.686	$p = .006$	D, F, S
201, 308	Driving, Fixed	23.686	$p < .001$	D, F, M, S
201, 318	Driving, Fixed	23.686	$p < .001$	D, F, M, S
201, 316	Driving, Fixed	22.771	$p < .001$	D, F, M, S
201, 307	Driving, Fixed	22.170	$p < .001$	D, F, L, M, S
201, 302	Driving, Fixed	19.212	$p = .001$	D, F, S
203, 306	Driving, Fixed	18.258	$p = .010$	D, F, M, S
203, 308	Driving, Fixed	17.562	$p = .018$	D, F, S
203, 318	Driving, Fixed	17.562	$p = .032$	F(V), S
203, 316	Driving, Fixed	16.646	$p = .030$	F(V), S
203, 307	Driving, Fixed	16.045	$p = .035$	D, F, L, S
204, 306	Driving, Fixed	10.199	$p = .040$	D, F, M, S
204, 308	Driving, Fixed	9.503	$p = .076$	D, F, MD, S
204, 316	Driving, Fixed	8.587	$p = .130$	D, F, MD, S
204, 307	Driving, Fixed	7.986	$p = .127$	D, F, MD, S
306, 312	Fixed, Rotary	12.918	$p = .003$	F, M, S
306, 310	Fixed, Rotary	12.798	$p = .042$	F, M, S
312, 318	Rotary, Fixed	12.222	$p = .044$	D, F, S
308, 310	Fixed, Rotary	12.102	$p = .080$	F, S
312, 316	Rotary, Fixed	11.306	$p = .008$	D, F, S
310, 316	Rotary, Fixed	11.186	$p = .132$	D, F, S
307, 312	Fixed, Rotary	10.705	$p = .005$	F, L, S
308, 312	Rotary, Rotary	12.222	$p = .005$	F, S
309, 312	Rotary, Rotary	9.164	$p = .016$	F

[1] D = Display Type, F = Field of View, L = System Latency, M = Motion Base, MD = Motion Base DOF, R = Resolution (driving sims only), S = Simulator Type

The equipment features for the significant study comparisons shown in Table 36 indicate that overall, display field-of-view differed for all of the significant comparisons on the Disorientation subscale. Although, only the vertical field-of-view was different for two of the study pairs (Study 203 and 316 and Study 203 and 318). Differences in the type of display (60%) and motion base (i.e., motion versus no motion; 58%) were present in many of the significant comparisons, whereas system latency differed in 27% of the study pairs. However,

differences in the degrees of freedom of the motion base was only noted in 3% of the significant comparisons and, because information on the display resolution was only available for the Driving simulators, this equipment feature only differed in four of the study pairs.

In terms of the individual study comparisons, the results in Table 35 indicate that the mean Disorientation score for Driving simulator Study 202 (47.54) was significantly greater than all of the other Driving simulators. Additionally, the Driving simulator in Study 201 had a significantly greater mean (27.17) than the Driving simulator in Study 204 (12.98). As shown in Table 36, the equipment differences in all of these significant Driving simulator comparisons included the type of display, field-of-view, and display resolution. Differences in the motion base were also noted in three of the significant study pairs (Studies 202 and 203, 202 and 204, and 201 and 204), whereas the overall system latency differed in two of the significant comparisons (Study 201 and 202 and Study 202 and 203).

The results of the other significant within system type comparisons revealed that there were no significant differences between the Fixed-Wing studies and there were only two significant Rotary-Wing study pairs. In these two comparisons, the mean Disorientation score for the Rotary-Wing simulator in Study 312 (15.70) was significantly greater than the mean score in Study 308 (3.48) and Study 309 (6.54). Table 36 shows that these two significant Rotary-Wing comparisons differed in terms of their field-of-view.

The results for the between system type comparisons shown in Table 35 revealed that the mean Disorientation score for the Driving simulator in Study 202 (47.54) was significantly greater than all of the Rotary- and Fixed-Wing simulators. Table 36 indicates that the equipment for Study 202 differed from every significant Rotary-Wing comparison on the field-of-view and motion base and differences in system latency occurred in 56% of the comparisons. Differences

in field-of-view were present in all of the Study 202 and Fixed-Wing comparisons whereas the motion base differed in 63% of the comparisons, and display type differed in 50%. However, differences in system latency was only noted in one of the significant comparisons (Study 202 and 307). Similar results were found for the Driving simulator in Study 201. Specifically, the mean score for Study 201 (27.17), which was the second highest mean out of all of the simulators, was significantly greater than all but one of the Rotary-Wing simulators (Study 310; mean = 15.58) as well as all but one of the Fixed-Wing simulators (Study 303; mean = 17.01). Differences in equipment for all of these significant study pairs included field-of-view and display type (cf. Table 36). Differences in system latency were also noted in 43% of the comparisons between Study 201 and the Rotary-Wing simulators, but only one study pair in the Fixed-Wing comparisons (Study 201 and 307). Additionally, motion base differences occurred in all of the significant comparisons between Study 201 and the Rotary-Wing simulators.

Other significant differences between Driving and flight simulators on the Disorientation subscale were shown for Studies 203 and 204. The mean for the Driving simulator in Study 203 (21.04) was significantly greater than the Rotary-Wing simulators in Studies 309 (6.54) and 315 (5.97). The mean in this study was also significantly greater than five of the Fixed-Wing studies (Studies 306, 307, 308, 316, and 318), which all had mean Disorientation scores that were less than, or in one case (Study 307), equal to 5.0 (i.e., negligible mean sickness scores). Similarly, the mean Disorientation score for the Driving simulator in Study 204 (12.98) was significantly greater than half of the Fixed-Wing studies (Studies 306, 307, 308, and 316). While the field-of-view differed for all of these significant comparisons, type of display differed for the two Rotary-Wing comparisons with Study 203 and all of the Fixed-Wing comparisons with Study 204, but only three of the five Fixed-Wing comparisons with Study 203 (cf. Table 36). Additionally,

differences in motion base were only noted in the Fixed-Wing Study 306 comparisons with Driving Studies 203 and 204, but differences in the degrees of freedom of the motion base occurred in the other Fixed-Wing comparisons with Study 204.

Lastly, the results for the Fixed- and Rotary-Wing study comparisons indicated that the Rotary-Wing simulator used in Study 312 had a significantly greater Disorientation mean (15.70) than the Fixed-Wing simulators in Studies 306 (2.78), 307 (5.00), 316 (4.40), and 318 (3.48). Likewise, the mean for the Rotary-Wing simulator in Study 310 (15.58) was significantly greater than the Fixed-Wing simulators in Studies 306 (2.78), 308 (3.48), and 316 (4.40). Differences in field-of-view were noted for all of the significant comparisons (cf. Table 36). Other noted equipment differences in the comparisons with Study 312 included display type (80% of the comparisons) and motion base (60%). Differences in display type were only present in two of the significant comparisons with Study 312 (Studies 316 and 318) and one of the Study 310 comparisons (Study 316). Similarly, motion base differences only occurred in two study pairs (Study 306 and 310 and Study 306 and 312) while system latency only differed in the Study 307 and 312 comparison.

SSQ Subscale Scores for VR Systems

In order to determine whether there were differences in sickness severity on the individual SSQ subscales between various types of VR system configurations, three One-Way ANOVAs were run on the subscale scores (one each for Nausea, Oculomotor, Disorientation) from 16 VR studies, where each study represented a homogenous set of equipment features. The results of the analysis on the Nausea subscale revealed a statistically significant difference among the discrete VR studies, $F(15, 1176) = 7.52$, $p < .001$. Table 37 presents the descriptive

statistics (i.e., means and standard deviations) of the Nausea subscale scores for each of the VR studies, which are grouped by type of VR system.

Table 37. Means and Standard Deviations of the SSQ Nausea Subscale Score for Studies Using Three Types of VR Systems

			Nausea Score	
VR System	Study Number	n	M	SD
HMD	101	47	16.24	18.11
	102	13	50.64	35.99
	103	25	19.08	30.29
	104	19	21.59	33.02
	105	81	25.91	30.33
	106	30	26.39	34.78
	107	200	20.46	25.35
	108	197	25.28	25.45
	109	194	27.39	27.65
	110	211	26.77	25.32
	111	32	6.26	7.51
	112	39	4.65	7.22
	113	12	3.97	6.38
BOOM	650	25	18.7	26.48
	651	32	8.35	8.99
CAVE	725	35	1.91	3.87
	Total	1192	22.2	26.13

Since the results of the ANOVA were significant, post hoc Games-Howell Tests were used to identify which study pairs had significant differences in their mean Nausea scores. Table 38 below provides a summary of the multiple comparison test results.

Table 38. Results of the Multiple Comparison Tests on the Nausea Subscale Scores for VR Studies

	Study	102	109	110	106	105	108	104	107	103	650	101	651	111	112	113	725
Study	Mean	50.64	27.39	26.77	26.39	25.91	25.28	21.59	20.46	19.08	18.70	16.24	8.35	6.26	4.65	3.97	1.91
102	50.64												*	*	*	*	*
109	27.39												*	*	*	*	*
110	26.77												*	*	*	*	*
106	26.39																*
105	25.91												*	*	*	*	*
108	25.28												*	*	*	*	*
104	21.59																
107	20.46												*	*	*	*	*
103	19.08																
650	18.70																
101	16.24														*	*	*
651	8.35																*
111	6.26																
112	4.65																
113	3.97																
725	1.91																

Table 39 provides a list of the equipment features that differed between each pair of studies shown above in Table 38, which were identified as statistically different on the Nausea subscale, and the associated significance level for each comparison. The significant study pairs in the table are ordered from largest to smallest mean difference.

Table 39. Equipment Differences and Significance Levels for VR Studies with Significantly Different Mean Nausea Scores

Study Pair	VR Type	Mean Difference	Significance	Equipment Differences[1]
102, 113	HMD, HMD	46.660	$p = .025$	R, S, W
102, 112	HMD, HMD	45.988	$p = .027$	F, R, S, W
102, 111	HMD, HMD	44.375	$p = .035$	F, R, S, W
109, 113	HMD, HMD	23.416	$p < .001$	TS
110, 113	HMD, HMD	22.791	$p < .001$	TS
109, 112	HMD, HMD	22.743	$p < .001$	TS, W
110, 112	HMD, HMD	22.119	$p < .001$	TS, W
105, 113	HMD, HMD	21.936	$p < .001$	F, I, R, TS, TL
108, 113	HMD, HMD	21.304	$p < .001$	TS
105, 112	HMD, HMD	21.263	$p < .001$	F, I, R, TS, TL
109, 111	HMD, HMD	21.130	$p < .001$	TS, W
108, 112	HMD, HMD	20.631	$p < .001$	TS, W
110, 111	HMD, HMD	20.506	$p < .001$	TS, W
105, 111	HMD, HMD	19.650	$p < .001$	F, I, R, TS, TL
108, 111	HMD, HMD	19.018	$p < .001$	TS, W
107, 113	HMD, HMD	16.488	$p < .001$	TS
107, 112	HMD, HMD	15.816	$p < .001$	TS, W
107, 111	HMD, HMD	14.203	$p < .001$	TS, W
101, 113	HMD, HMD	12.263	$p = .029$	F, I, S, TS, W
101, 112	HMD, HMD	11.591	$p = .014$	F, I, S, TS, W
102, 651	HMD, BOOM	42.288	$p = .049$	D, F, R, T
109, 651	HMD, BOOM	19.043	$p < .001$	D, F, R, T
110, 651	HMD, BOOM	18.419	$p < .001$	D, F, R, T
105, 651	HMD, BOOM	17.564	$p = .001$	D, F, R, T
108, 651	HMD, BOOM	16.931	$p < .001$	D, F, R, T
107, 651	HMD, BOOM	12.116	$p < .001$	D, F, R, T
102, 725	HMD, CAVE	48.727	$p = .018$	D, F, TS, W
109, 725	HMD, CAVE	25.483	$p < .001$	D, F, I, TS, W
110, 725	HMD, CAVE	24.858	$p < .001$	D, F, I, TS, W

106, 725	HMD, CAVE	24.486	$p = .039$	D, F, T
105, 725	HMD, CAVE	24.003	$p < .001$	D, F, TL, TS
108, 725	HMD, CAVE	23.371	$p < .001$	D, F, I, TS, W
107, 725	HMD, CAVE	18.555	$p < .001$	D, F, I, TS, W
101, 725	HMD, CAVE	14.330	$p < .001$	D, F, TS, W
651, 725	BOOM, CAVE	6.439	$p = .038$	D, F, T

[1] D = Display Type, F = Field-of-View, L = System Latency, M = Motion Base, MD = Motion Base DOF
R = Resolution (driving sims only), S = Simulator Type

The equipment features for the significant study comparisons shown in Table 39 indicate that overall, differences related to the head tracker speed were present in the largest amount of the study pairs (69%) followed by differences in the field-of-view (63%), weight of the display (54%), and display resolution (34%). Additionally, IPD adjustability and head tracking differences accounted for equipment differences in 26% and 23% of the significant comparisons between studies. On the other hand, differences in screen size (14%) and head tracker latency (11%) were only noted in a few of the significant study comparisons.

In terms of the individual study comparisons, the results in Table 38 revealed that the only significant within VR system type comparisons were with the HMD systems in which 26% of the study pairs had significantly different mean Nausea scores. The results for these comparisons indicated that the mean scores for Studies 102 (50.64), 105 (25.91), 107 (20.46), 108 (25.28), 109 (27.39), and 110 (26.77) were each significantly greater than the mean scores for the HMDs in Studies 111 (6.26), 112 (4.65), and 113 (3.97). Similarly, the mean for Study 101 (16.24) was also significantly greater than Studies 112 and 113. The equipment features in Table 39 show that in these significant study comparisons, differences in the speed of the head tracker were present in most (85%) of the study pairs while the weight of the display also differed in many (65%) of the comparisons. A smaller proportion of the significant HMD study

pairs had differences in field-of-view (35%), display resolution (30%), IPD adjustability (25%), screen size (25%), and head tracker latency (15%).

The results for the between VR system comparisons indicated that the mean Nausea score for the BOOM system in Study 651 (8.35) was significantly lower than six of the HMD studies (Studies 102, 105, 107, 108, 109, and 110), but significantly greater than the mean for the CAVE study (1.91). Conversely, the mean for the BOOM system in Study 650 (18.70) was not significantly different than any of the HMD studies nor the CAVE study. Table 39 shows that the equipment features that differed between all of the significant BOOM and HMD studies were the field-of-view, display resolution, use of a head tracker, and of course, the type of display. Difference in field-of-view, use of a head tracker, and display type were also noted in the significant BOOM and CAVE study pair.

Finally, comparisons between the HMD and CAVE systems also showed significant differences in the Nausea score. These results indicated that the mean for the CAVE system (1.91) was significantly lower than 62% (8 out of 13) of the HMD studies. For these significant comparisons, field-of-view and display type differed in all of the study pairs and differences in head tracker speed were present in all but one of the study pairs (Study 106 and 725), where there was a difference in whether head tracking was provided. The other equipment features that differed in many of these significant HMD-CAVE comparisons was the weight of the display (75%) and IPD adjustability (50%).

The next ANOVA was conducted on the Oculomotor subscale scores for all of the VR studies. The results of the analysis revealed a statistically significant difference among the 16 VR studies, $F(15, 1176) = 7.14$, $p < .001$. In Table 40, the Oculomotor subscale score means and standard deviations, grouped by type of VR system, are shown for each of the simulator studies.

Table 40. Means and Standard Deviations of the SSQ Oculomotor Subscale Score for Studies Using Three Types of VR Systems

			Oculomotor Score	
VR System	Study Number	n	M	SD
HMD	101	47	15.64	13.26
	102	13	36.15	22.13
	103	25	19.4	18.32
	104	19	18.75	21.20
	105	81	29.38	23.53
	106	30	24.26	17.63
	107	200	18.42	20.64
	108	197	22.09	19.37
	109	194	24.97	19.69
	110	211	28.99	22.10
	111	32	10.66	16.66
	112	39	10.5	10.08
	113	12	13.9	10.14
BOOM	650	25	23.65	22.46
	651	32	20.85	16.45
CAVE	725	35	7.58	6.37
	Total	1192	22.31	20.40

Post hoc Games-Howell Tests were then used to identify which study pairs had significant differences in their mean Oculomotor scores. Table 41 below provides a summary of the multiple comparison test results.

Table 41. Results of the Multiple Comparison Tests on the Oculomotor Subscale Scores for VR Studies

	Study	102	105	110	109	106	650	108	651	103	104	107	101	113	111	112	725
Study	Mean	36.15	29.38	28.99	24.97	24.26	23.65	22.09	20.85	19.40	18.75	18.42	15.64	13.90	10.66	10.50	7.58
102	36.15																*
105	29.38											*	*	*	*	*	*
110	28.99											*	*	*	*	*	*
109	24.97												*		*	*	*
106	24.26															*	*
650	23.65																
108	22.09															*	*
651	20.85																*
103	19.40																
104	18.75																
107	18.42															*	*
101	15.64																*
113	13.90																
111	10.66																
112	10.50																
725	7.58																

Table 42 provides a list of the equipment features that differed between each pair of studies shown above in Table 41, which were identified as statistically different on the Oculomotor subscale, and the associated significance level for each comparison. The significant study pairs in the table are ordered from largest to smallest mean difference.

Table 42. Equipment Differences and Significance Levels for VR Studies with Significantly Different Mean Oculomotor Scores

Study Pair	VR Type	Mean Difference	Significance	Equipment Differences[1]
105, 112	HMD, HMD	18.889	$p < .001$	F, I, R, TS, TL
105, 111	HMD, HMD	18.725	$p = .001$	F, I, R, TS, TL
110, 112	HMD, HMD	18.495	$p < .001$	TS, W
110, 111	HMD, HMD	18.331	$p < .001$	TS, W
105, 113	HMD, HMD	15.488	$p = .028$	F, I, R, TS, TL
110, 113	HMD, HMD	15.094	$p = .015$	TS
109, 112	HMD, HMD	14.472	$p < .001$	TS, W
109, 111	HMD, HMD	14.308	$p = .006$	TS, W
106, 112	HMD, HMD	13.761	$p = .031$	F, I, R, T
101, 105	HMD, HMD	13.740	$p = .004$	F, R, TS
101, 110	HMD, HMD	13.347	$p < .001$	F, I, S, W
108, 112	HMD, HMD	11.591	$p < .001$	TS, W
105, 107	HMD, HMD	10.965	$p = .030$	F, I, R, TS
107, 110	HMD, HMD	10.571	$p < .001$	---
101, 109	HMD, HMD	9.323	$p = .016$	F, I, S, W
107, 112	HMD, HMD	7.924	$p = .033$	TS, W
102, 725	HMD, CAVE	28.571	$p = .025$	D, F, TS, W
105, 725	HMD, CAVE	21.804	$p < .001$	D, F, TL, TS
110, 725	HMD, CAVE	21.411	$p < .001$	D, F, I, TS, W
109, 725	HMD, CAVE	17.387	$p < .001$	D, F, I, TS, W
106, 725	HMD, CAVE	16.676	$p = .002$	D, F, T
108, 725	HMD, CAVE	14.506	$p < .001$	D, F, I, TS, W
107, 725	HMD, CAVE	10.839	$p < .001$	D, F, I, TS, W
101, 725	HMD, CAVE	8.064	$p = .039$	D, F, TS, W
651, 725	BOOM, CAVE	13.265	$p = .010$	D, F, T

[1] D = Display Type, F = Field-of-View, L = System Latency, M = Motion Base, MD = Motion Base DOF
R = Resolution (driving sims only), S = Simulator Type

The equipment features for the significant study comparisons shown in Table 42 indicate that overall, differences related to the head tracker speed were present in the largest amount of the study pairs (76%) followed by differences in the field-of-view (68%), weight of the display (56%), and IPD adjustability (44%). Additionally, display resolution differences accounted for equipment differences in 24% of the significant comparisons between studies. However, differences in head tracker latency (16%), use of a head tracker (12%), and screen size (8%) and were only noted in a few of the significant study comparisons.

For the individual study comparisons, the results in Table 41 revealed that, like the Nausea score, the only significant within VR system type comparisons were with the HMD systems in which 21% of the study pairs had significantly different mean Oculomotor scores. The results for these comparisons indicated that the mean scores for Studies 105 (29.38) and 110 (28.99) were both significantly greater than the mean scores for the HMDs in Studies 101 (15.64), 107 (18.42), 111 (10.66), 112 (10.50), and 113 (13.90). Similarly, the mean for Study 109 (24.97) was also significantly greater than Studies 101, 111 and 112. Studies 106 (24.26), 107 (18.42), and 108 (22.09) also had significantly greater means than Study 112. The equipment features in Table 42 show that in these significant study comparisons, differences in the speed of the head tracker were present in most (75%) of the study pairs while the weight of the display and field-of-view also differed in many (both 50%) of the comparisons. The other equipment differences that were present in these significant comparisons were the IPD adjustability (44%), display resolution (38%), and head tracker latency (19%). However, as shown in Table 42, there were no equipment differences noted between Study 107 and 110, which will be addressed in Chapter 5.

The results for the between VR system comparisons indicated that the mean Oculomotor score for the CAVE study (7.58) was significantly lower than more than half (62%) of the HMD studies (Studies 101, 102, 105, 106, 107, 108, 109, and 110). For these significant comparisons, field-of-view and display type differed in all of the study pairs and differences in head tracker speed were present in all but one of the study pairs (Study 106 and 725), where there was a difference in whether head tracking was provided. The other equipment features that differed in many of these significant HMD-CAVE comparisons was the weight of the display (75%) and IPD adjustability (50%).

Finally, the results for the comparisons with the BOOM system indicated that there were no significant within system type differences (i.e., Study 650 and 651) and neither of the BOOM studies were significantly different from any of the HMD studies. The comparisons with the CAVE study, however, revealed that the BOOM in Study 651 had a significantly greater mean Oculomotor score (20.85) than the CAVE system (7.58). The equipment features in these two studies differed in terms of their field-of view, whether head tracking was provided, and of course, the type of display.

The last analysis was a One-Way ANOVA was run on the Disorientation subscale scores from the 16 VR studies. The results of the analysis revealed a statistically significant difference among the discrete simulator studies, $F(15, 1176) = 5.12$, $p < .001$. The descriptive statistics (i.e., means and standard deviations) of the Disorientation subscale scores for each of the VR studies, grouped by type of VR system, are presented in Table 43.

Table 43. Means and Standard Deviations of the SSQ Disorientation Subscale Score for Studies Using Three Types of VR Systems

			Disorientation Score	
VR System	Study Number	n	M	SD
HMD	101	47	19.25	21.04
	102	13	54.61	39.97
	103	25	24.5	29.33
	104	19	26.37	48.86
	105	81	42.45	58.27
	106	30	26.91	35.62
	107	200	30.48	34.76
	108	197	33.14	36.81
	109	194	36.81	36.69
	110	211	35.23	34.72
	111	32	13.92	11.73
	112	39	12.14	11.59
	113	12	4.64	6.85
BOOM	650	25	28.4	43.55
	651	32	13.48	13.46
CAVE	725	35	7.95	8.46
	Total	1192	30.68	36.56

Games-Howell Tests were then used for the post hoc multiple comparisons to identify which study pairs had significant differences in their mean Disorientation scores. Table 44 below provides a summary of the multiple comparison test results.

Table 44. Results of the Multiple Comparison Tests on the Disorientation Subscale Scores for VR Studies

	Study	102	105	109	110	108	107	650	106	104	103	101	111	651	112	725	113
Study	Mean	54.61	42.45	36.81	35.23	33.14	30.48	28.40	26.91	26.37	24.50	19.25	13.92	13.48	12.14	7.95	4.64
102	54.61																*
105	42.45												*	*	*	*	*
109	36.81											*	*	*	*	*	*
110	35.23											*	*	*	*	*	*
108	33.14												*	*	*	*	*
107	30.48												*	*	*	*	*
650	28.40																
106	26.91																
104	26.37																
103	24.50																
101	19.25																*
111	13.92																
651	13.48																
112	12.14																
725	7.95																
113	4.64																

Table 45 provides a list of the equipment features that differed between each pair of studies shown above in Table 44, which were identified as statistically different on the Oculomotor subscale, and the associated significance level for each comparison. The significant study pairs in the table are ordered from largest to smallest mean difference.

Table 45. Equipment Differences and Significance Levels for VR Studies with Significantly Different Mean Disorientation Scores

Study Pair	VR Type	Mean Difference	Significance	Equipment Differences[1]
102, 113	HMD, HMD	49.969	$p = .032$	R, S, W
105, 113	HMD, HMD	37.807	$p < .001$	F, I, R, TS, TL
109, 113	HMD, HMD	32.169	$p < .001$	TS
110, 113	HMD, HMD	30.589	$p < .001$	TS
105, 111	HMD, HMD	28.527	$p = .006$	F, I, R, TS, TL
105, 112	HMD, HMD	28.527	$p = .006$	F, I, R, TS, TL
108, 113	HMD, HMD	28.499	$p < .001$	TS
107, 113	HMD, HMD	25.845	$p < .001$	TS
109, 112	HMD, HMD	24.889	$p < .001$	TS, W
110, 112	HMD, HMD	23.093	$p < .001$	TS, W
109, 111	HMD, HMD	22.889	$p < .001$	TS, W
110, 111	HMD, HMD	21.309	$p < .001$	TS, W
108, 112	HMD, HMD	21.004	$p < .001$	TS, W
108, 111	HMD, HMD	19.219	$p < .001$	TS, W
107, 112	HMD, HMD	18.349	$p < .001$	TS, W
101, 109	HMD, HMD	17.558	$p = .003$	F, I, S, W
107, 111	HMD, HMD	16.565	$p < .001$	TS, W
101, 110	HMD, HMD	15.978	$p = .007$	F, I, S, W
101, 113	HMD, HMD	14.611	$p = .016$	F, I, S, TS, W
105, 651	HMD, BOOM	28.962	$p = .006$	D, F, R, T
109, 651	HMD, BOOM	23.324	$p < .001$	D, F, R, T
110, 651	HMD, BOOM	21.744	$p < .001$	D, F, R, T
108, 651	HMD, BOOM	19.654	$p < .001$	D, F, R, T
107, 651	HMD, BOOM	17.000	$p < .001$	D, F, R, T
105, 725	HMD, CAVE	34.493	$p < .001$	D, F, TL, TS
109, 725	HMD, CAVE	28.855	$p < .001$	D, F, I, TS, W
110, 725	HMD, CAVE	27.275	$p < .001$	D, F, I, TS, W
108, 725	HMD, CAVE	25.185	$p < .001$	D, F, I, TS, W
107, 725	HMD, CAVE	22.531	$p < .001$	D, F, I, TS, W

[1] D = Display Type, F = Field-of-View, L = System Latency, M = Motion Base, MD = Motion Base DOF
R = Resolution (driving sims only), S = Simulator Type

The equipment features for the significant study comparisons shown in Table 45 indicate that overall, differences related to the head tracker speed were present in the largest amount of the study pairs (72%) followed by differences in the field-of-view and weight of the display (both 55%), IPD adjustability (34%), and display resolution (31%). Although, differences in the use of a head tracker (17%), head tracker latency (14%), and screen size (14%) and were only noted in a few of the significant study comparisons.

The results shown in Table 44 for the individual study comparisons once again revealed that the only significant within VR system type comparisons were with the HMD systems in which 24% of the study pairs had significantly different mean Disorientation scores. The results for these comparisons indicated that the mean scores for Study 113 (4.64) was significantly lower than the mean scores for the HMDs in Studies 101 (19.25), 102 (54.61), 105 (42.45), 107 (30.48), 108 (33.14), 109 (36.81), and 110 (35.23). Similarly, the mean for Studies 111 (13.92) and 112 (12.14) were also significantly lower than Studies 105, 107, 108, 109, and 110. Additionally, Study 101 (19.25) had a significantly lower mean than Studies 109 and 110. Table 45 shows that in these significant study comparisons, differences in the speed of the head tracker were present in most (84%) of the study pairs while the weight of the display also differed in many (63%) of the comparisons. A smaller proportion of the significant HMD study pairs had differences in field-of-view (32%), IPD adjustability (32%), display resolution (21%), screen size (21%), and head tracker latency (16%).

The results for the between VR system comparisons indicated that the mean Disorientation score for the BOOM system in Study 651 (13.48) was significantly lower than five of the HMD studies (Studies 105, 107, 108, 109, and 110), but not significantly different than the mean for the CAVE study (7.95). The mean for the BOOM system in Study 650 (18.70)

was also not significantly different than any of the HMD studies nor the CAVE study. Table 45 shows that the equipment features that differed between all of the significant BOOM and HMD studies were the field-of-view, display resolution, use of a head tracker, and the type of display.

Finally, comparisons between the HMD and CAVE systems also showed significant differences in the Disorientation score. These results indicated that the mean for the CAVE system (7.95) was significantly lower than 40% (5 out of 13) of the HMD studies. For these significant comparisons, field-of-view, display type, and head tracker speed differed in all of the study pairs. In all but one of the study pairs, differences in IPD adjustability (Study 105 and 725) and weight of the display (also Study 105 and 725) were noted differences in the equipment. Conversely, a difference in the latency of the head tracker was only present in one of the significant HMD-CAVE comparisons (Study 105 and 725).

Profile Validation

Two datasets (one from a simulator and one from a VR system) that were *not* included in the original database were used to validate the findings of the profile analyses that were derived from the preceding analyses. Although there was not enough information to validate the results for specific engineering characteristics of the systems, the studies from the original database were matched on certain aspects of the new datasets in both of the analyses described below.

Simulator Validation for Proportional Subscale Scores

The previous analyses on the profiles for simulator type indicated that Driving simulators were not significantly different than Fixed- and Rotary-Wing simulators on the Nausea subscale, but they were different on the Oculomotor and Disorientation subscales. These analyses also

showed that there were no differences between Fixed- and Rotary-Wing simulators on the Nausea and Oculomotor subscales, but they were different on the Disorientation subscale. Therefore, since the new simulator dataset was from a Rotary-Wing study, the studies from the original database that were used to validate the profiles for the new simulator study data were also from Rotary-Wing studies (Studies 305, 309, 310, 311, 312, 313, 314, 315, and 317). A multivariate analysis of variance (MANOVA) was conducted to assess if there were profile differences between the two Rotary-Wing simulator datasets. The results revealed there was not a statistically significant difference between the means for the two simulator datasets, Pillai's Trace = .001, $F(2, 512) = 0.001$, $p = .721$, which indicates that the two profiles are similar. The means and standard deviations of the proportional subscale scores for the two Rotary-Wing simulator datasets are provided in Table 46.

Table 46. Means and Standard Deviations Comparing Rotary-Wing Simulators in the Original Database and the New Rotary-Wing Simulator Study

		Proportional N		Proportional O		Proportional D	
Data Source	n	M	SD	M	SD	M	SD
Original Database	496	0.293	0.291	0.532	0.332	0.175	0.234
New Study	19	0.242	0.332	0.558	0.378	0.201	0.232
Total	515	0.291	0.292	0.533	0.333	0.176	0.234

The results of this analysis was somewhat anticipated based on the previous profile analyses for simulator studies. Specifically, the previous results for Rotary-Wing simulator study comparisons indicated that there were no significant differences on the Nausea or Oculomotor subscale scores and only one pair of studies that differed on the Disorientation subscale. However, the significance level for this comparison ($p = .065$) suggested that the difference could have been a spurious result.

VR Validation for Proportional Subscale Scores

The previous analyses on the profiles for VR type indicated that HMDs were significantly different than CAVE systems on the Nausea and Oculomotor subscales, but they were not different on the Disorientation subscale. These analyses also showed that HMDs were not significantly different than BOOM systems on the Nausea and Disorientation subscales, but they were different on the Oculomotor subscale. Therefore, since the new VR dataset was from an HMD study, the studies from the original database that were used to validate the profiles for the new HMD study data were also HMD studies (Studies 101-113). A multivariate analysis of variance (MANOVA) was conducted to assess if there were profile differences between the two HMD datasets. A statistically significant difference was found between the means for the two HMD datasets, Pillai's Trace = .019, $F(2, 1116) = 10.60$, $p < .001$, which indicates that the two profiles are different. The means and standard deviations of the proportional subscale scores for the two HMD datasets are provided in Table 47.

Table 47. Means and Standard Deviations Comparing HMD Systems in the Original Database and the New HMD Study

		Proportional N		Proportional O		Proportional D	
Data Source	n	M	SD	M	SD	M	SD
Original Database	1100	0.281	0.245	0.368	0.278	0.351	0.264
New Study	19	0.026	0.063	0.470	0.335	0.505	0.330
Total	1119	0.277	0.245	0.370	0.279	0.353	0.265

Follow up Univariate ANOVAs indicated that two of the proportional subscale scores were significantly different for the two HMD datasets: $F(1, 1117) = 20.64$, $p < .001$ for the proportional Nausea subscale score and $F(1, 1117) = 6.32$, $p = .012$ for the proportional

Disorientation subscale score. However, the means for the proportional Oculomotor subscale score were not significantly different, $F(1, 1117) = 2.47$, $p = .116$. As shown in Table 47 above, the mean for the HMDs in the original dataset was greater than the HMD in the new dataset for the proportional Nausea subscale score, whereas the mean for the new dataset was greater than the original dataset for the proportional Disorientation subscale score.

The results of this analysis was also somewhat anticipated based on the previous profile analyses for VR studies, although the specific subscales that differed in this analysis were not anticipated. Specifically, the previous results for the HMD study comparisons indicated that the HMD study means differed for the Oculomotor and Disorientation subscales, but there were no significant differences on the Nausea subscale.

CHAPTER FIVE: DISCUSSION AND CONCLUSIONS

There were two primary objectives of the current research. One of the objectives was to quantitatively determine whether the sickness produced by exposure to simulators and VR devices were different. The other objective was to determine whether there were quantitative differences in the patterns of symptoms (i.e., the SSQ profiles) over diverse VE systems. Additionally, this research sought to determine the form of the relationship between different engineering features of the VE systems and the sickness symptoms produced as a result of exposure to them.

In order to accomplish these objectives, several different types of statistical analyses were conducted on a large database that contained SSQ data from a total of 2100 participants. These data represented sickness symptoms reported by individuals following exposure to six different types of VE systems (three types of simulators and three types of VR systems). A discussion of the results that were conducted to support the research objectives are provided below. First, a discussion of the results for the analyses regarding differences between different types of VE systems are presented. Then, the results for the analyses on the individual VE studies that were conducted to identify differences in SSQ profiles and symptom severity within and between the various VE systems are discussed along with the findings related to the engineering features of the systems.

Comparison of Symptom Profiles Between and Within Different Types of VE Systems

Several analyses were conducted in order to determine whether quantitative differences in SSQ profiles existed between and within different types of VE systems. In the subsequent sections, the results for the differences in profiles between simulator and VR systems are discussed first. Then, a discussion of the differences within the three types of simulator systems and the three types of VR systems are presented.

Profile Comparison for Simulator and VR Systems

In the literature on VE sickness, the terms cybersickness or virtual reality sickness have been commonly used to refer to the adverse effects produced by VR devices in order to distinguish the symptoms from those produced by simulators. A fundamental question that has not been previously addressed, however, is whether simulator sickness and cybersickness produce sufficiently different types of symptoms to justify the use of two separate terms. The results of the MANOVA and post hoc tests revealed a statistically significant difference between the two types of VE systems on all three of the proportional subscale scores. Figure 5 presents the profiles, based on the mean proportional subscales, for simulator and VR systems.

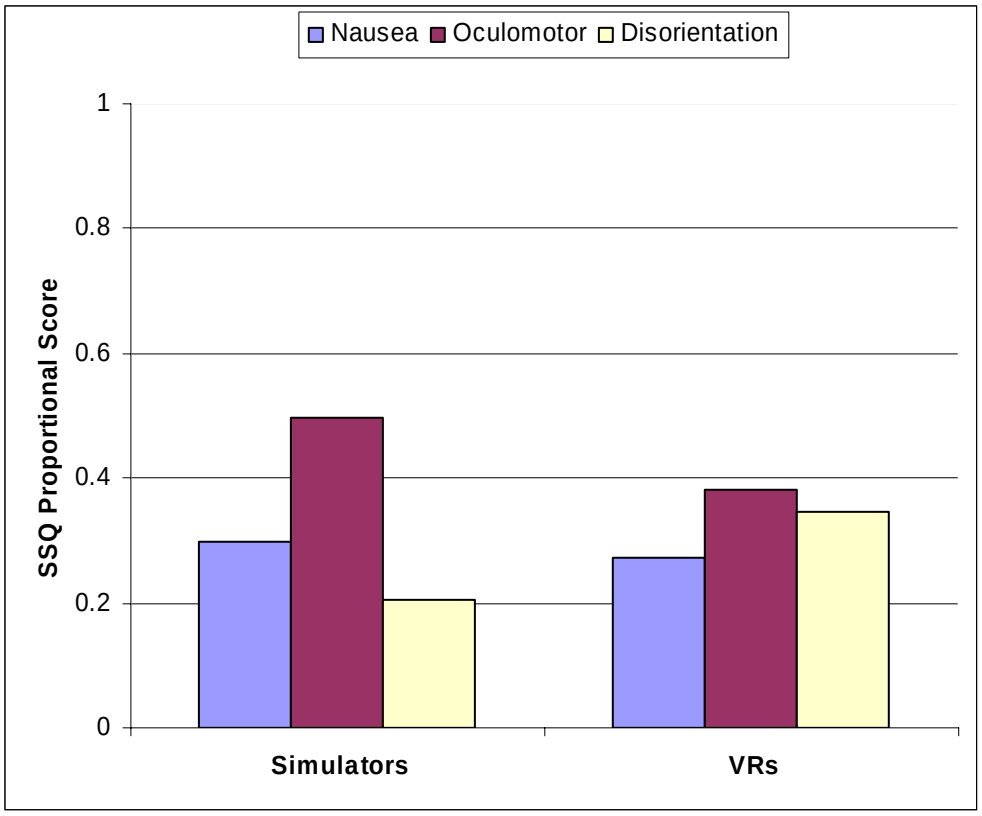

Figure 5. SSQ profiles for simulator and VR systems.

A visual comparison of the profiles shown in Figure 5 illustrates the statistical difference found between profiles for the two types of systems. The profile for simulators shows that Oculomotor discomfort produces the largest relative contribution to sickness followed by Nausea and Disorientation. The VR profile also exhibits a higher relative contribution of Oculomotor symptoms, but in this profile, Disorientation symptoms contribute more to sickness than the Nausea component.

In addition to profile differences, the results of the analysis on the SSQ Total Severity score also showed a difference between simulator and VR systems which indicated that the overall severity of sickness associated with exposure to VR systems was greater than simulator exposures. Taken together, these results provide quantitative evidence of a difference in sickness between the two types of VE systems and indicate that they represent distinct motion sickness

constructs. Accordingly, these results support the use of two separate terms (i.e., simulator sickness and cybersickness) to refer to the negative side-effects of exposure to these systems.

Profile Comparison for Three Types of Simulator Systems

The results of a separate MANOVA revealed a statistically significant difference within the three types of simulators (i.e., Fixed- and Rotary-Wing flight simulators and Driving simulators). The post hoc analyses indicated that although the three types of simulators did not differ on the proportional Nausea subscale, there were significant differences on the proportional Oculomotor subscale between the flight simulators and the Driving simulator and differences between all three of the simulators on the Disorientation subscale. The profiles for the three types of simulators, based on the mean proportional subscales, are shown in Figure 6.

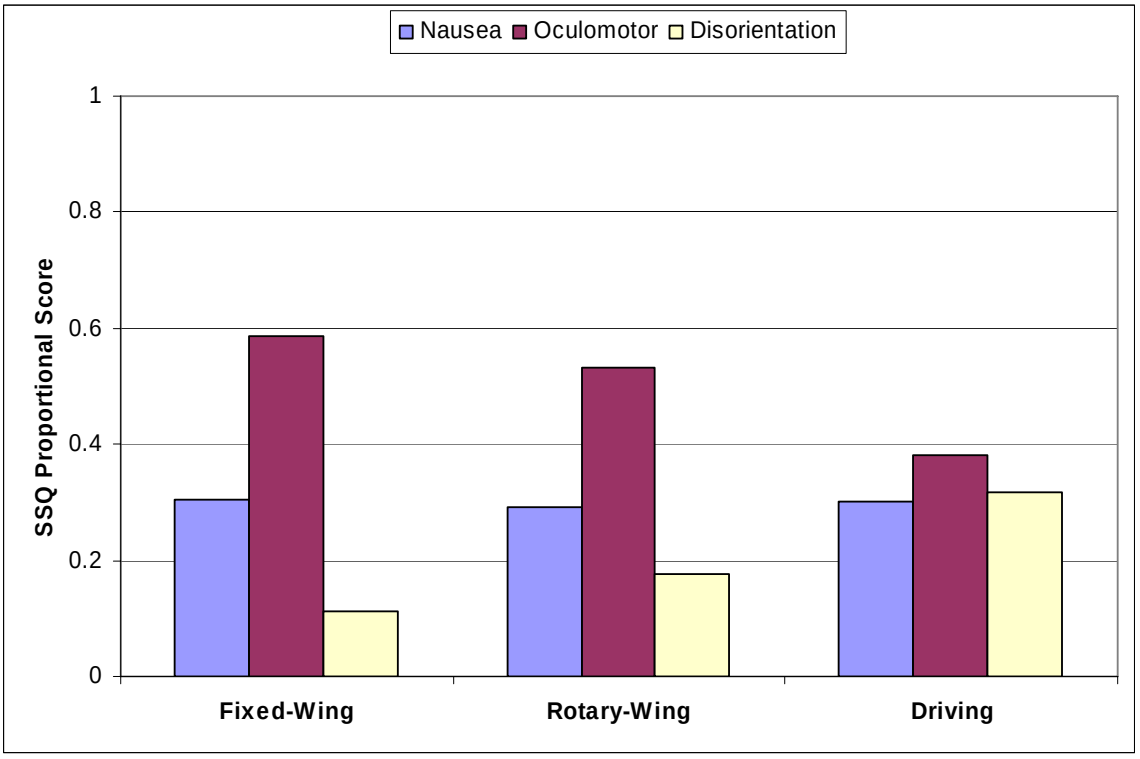

Figure 6. SSQ profiles for three types of simulator systems.

The overall profile is similar for the two flight simulators, which are also consistent with the average profile for simulators shown previously in Figure 5 (i.e., O>N>D). However, the Driving simulator has a different profile (O>D>N). The data in Figure 6 provide a visual confirmation of the statistical differences in profiles for the three simulators. Specifically, each of the profiles shows a similar level of contribution for the Nausea subscale. In contrast, the flight simulators have a fairly similar Oculomotor component compared to the Driving simulator which has a much smaller relative contribution of visual symptoms. A comparison of the profiles also reflects the differences in the relative contribution of Disorientation symptoms between all three of the simulator types; Driving simulators have the highest contribution of Disorientation followed by Rotary-Wing simulators and then Fixed-Wing simulators.

The analysis on the SSQ Total Severity score also showed significant differences within the three types of simulators which indicated that the overall severity of sickness was greatest in the Fixed-Wing flight simulators followed by Rotary-Wing flight simulators and then Driving simulators. Thus, these results also provide some quantitative evidence that there are differences in the sickness profiles within different types of simulator systems.

Profile Comparison for Three Types of VR Systems

The results of another MANOVA also revealed a statistically significant difference within the three types of VR systems (i.e., HMD, BOOM, and CAVE). The post hoc analyses, however, indicated that there were no differences in the mean proportional scores for the BOOM and CAVE systems, but there were differences between these two types of systems and the HMD systems. Figure 7 provides the profile, based on the mean proportional subscales, for each of the three VR systems.

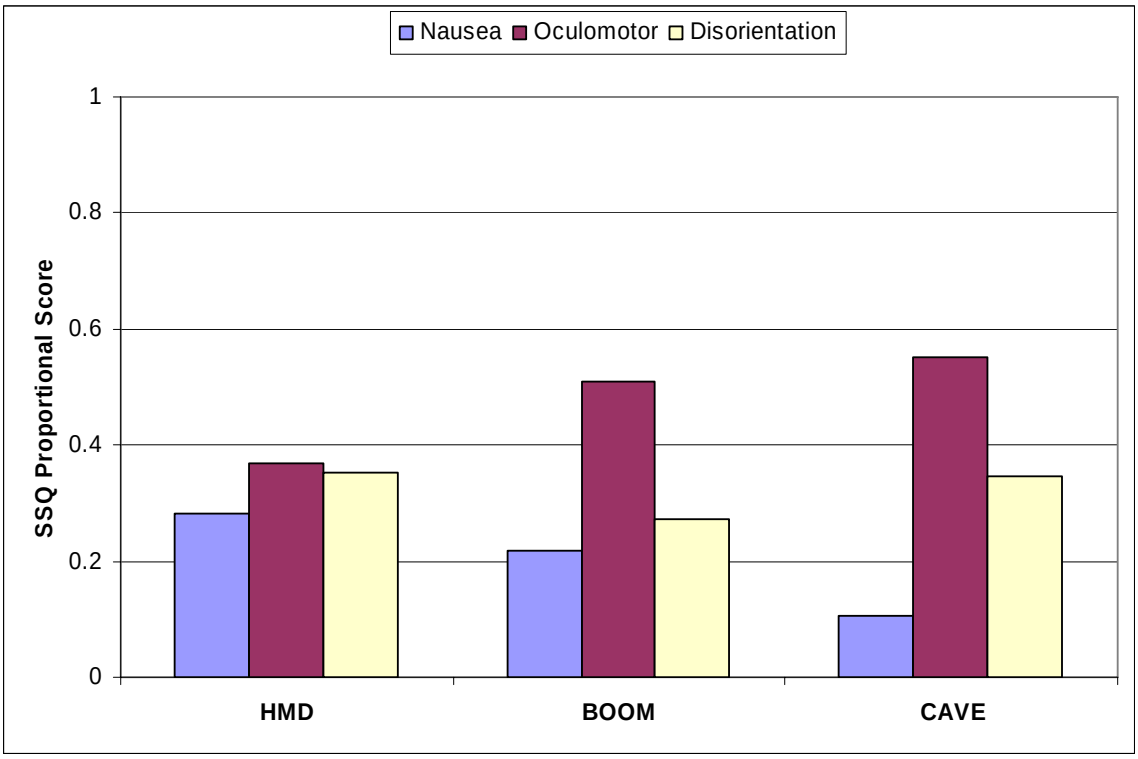

Figure 7. SSQ profiles for three types of VR systems.

The data in Figure 7 show that, as with the different types of simulators, the overall profile for the three types of VR systems is consistent with the average profile for VR systems shown previously in Figure 5 (i.e., O>D>N). However, Figure 7 also shows that there are obvious differences between the profiles, especially for the HMD profile. The profile for the HMD has a significantly higher proportional contribution of Nausea-type symptoms than the CAVE system. Conversely, the relative contribution of Oculomotor symptoms to the sickness reported was significantly lower in HMD systems than in both the BOOM and CAVE systems, which had similar proportional Oculomotor subscale scores. Although not readily apparent in the profiles, the results of the Total Severity score analysis also indicated that the overall severity of sickness was significantly higher for the HMD than the BOOM systems, which was significantly higher than the CAVE system. Therefore, these results indicate that the BOOM and

CAVE systems have similar profiles, which are different than the profile for HMD systems and the overall severity of sickness is worse in HMD systems than in the other two types of VR systems.

Comparison of Symptom Profiles and Sickness Severity for Individual VE Studies

The results for the analyses on the individual VE studies that were conducted to identify whether there were quantitative differences in SSQ profiles and symptom severity within and between the various VE systems are discussed below along with the findings related to the engineering features of the systems. In the subsequent sections, the results for the 21 individual simulator studies are discussed first. Then, a discussion of the results for the 16 individual VR studies is presented.

Profile and Sickness Severity Comparison for Simulator Studies

The results of the MANOVA for the proportional subscales scores revealed a statistically significant difference among the discrete simulator studies. Similarly, the One-Way ANOVAs for the SSQ Total Severity score and the subscale scores also indicated that there were statistically significant differences among the studies. A discussion of the post hoc multiple comparison tests for all of these analyses are provided below; the within system results discussion is presented first followed by the between system results.

Driving Simulators

The results of the multiple comparison tests for the proportional scores indicated that there were no significant differences for any of the Driving simulator studies on the Nausea subscale. In contrast, Studies 201 and 204 both had significant differences on the Oculomotor and Disorientation subscales while Studies 202 and 204 differed only on the Disorientation subscale. Therefore, two of the study pairs had similar profiles and two had different profiles. The profiles for each of the Driving simulator studies is shown in Figure 8 below.

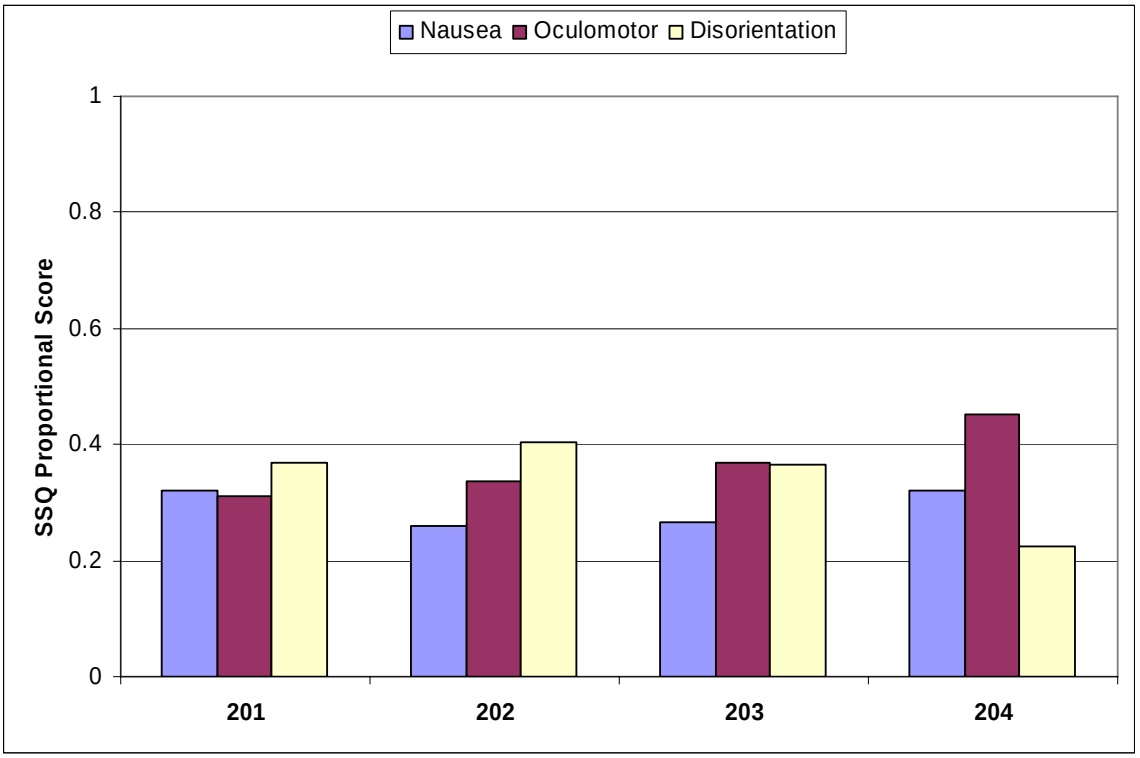

Figure 8. SSQ profiles for the Driving simulator studies.

As shown in Figure 8, the Driving simulator profiles that differed were Studies 201 (D>N>O) and 204 (O>N>D) and Studies 202 (D>O>N) and 204 (O>N>D). Additionally, a comparison of the individual study profiles in Figure 8 and the profile for the mean Driving simulator shown previously in Figure 6 reveal that only one of the studies (Study 203) has a

similar profile (i.e., O>D>N). Thus, the profile of the mean Driving simulator does not appear to provide an accurate reflection of the individual study profiles. Since SSQ data was only available for four Driving simulator studies, additional studies could potentially provide a more representative mean profile.

As with the differences in profiles between Study 202 and 204, there were also significant differences for this study pair on all of the SSQ subscale severity scores (Total Severity, Nausea [N], Oculomotor [O], and Disorientation [D]), in which the means for Study 202 were greater than Study 204. The other study pairs that showed significant differences in sickness severity on at least two of the subscales, but did not show differences on the proportional scores were Studies 202 and 203 and Studies 201 and 202. In the first study pair, Study 202 had a significantly higher mean on all three of the subscales (N, O, and D) compared to Study 203 whereas in the second study pair, the mean for Study 202 was significantly greater than Study 201 on the Oculomotor and Disorientation subscales.

Taken together, these results also indicate that there are quantitative differences in the SSQ profiles and symptom severity within the various types of Driving simulators.

Fixed-Wing Simulators

The results of the multiple comparison tests for the proportional scores indicated that there were no significant differences for any of the Fixed-Wing studies on the Disorientation subscale, one study pair (Study 303 and 307) differed only on the Oculomotor subscale and one study pair differed on both the Nausea and Oculomotor subscales (Study 302 and 307). The profiles for each of the Fixed-Wing simulator studies are shown in Figure 9 below.

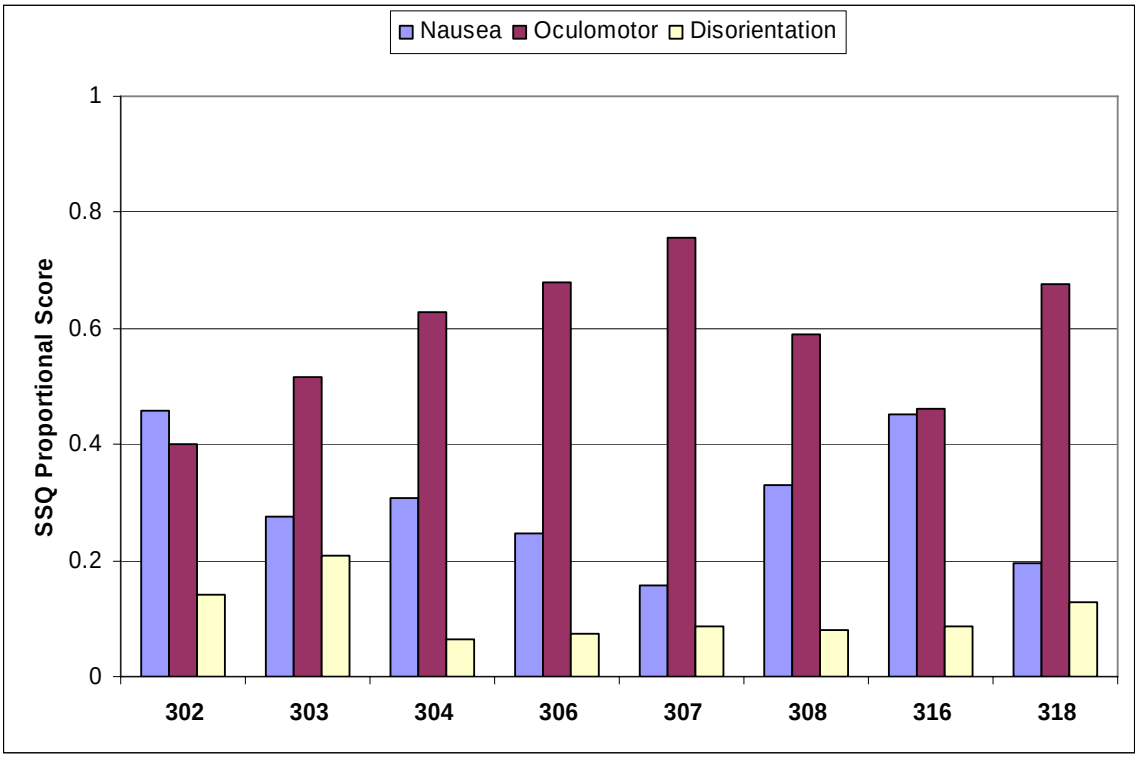

Figure 9. SSQ profiles for Fixed-Wing simulator studies.

A comparison of the profiles shown in Figure 9, indicate that Study 302 (N>O>D) appears to have a different profile than the other Fixed-Wing simulators (O>N>D). However, the analyses only showed a statistically significant difference between the profiles for Studies 302 and 307. The results for the other significant study pair (Study 303 and 307) only differed on one of the subscales. As shown in Figure 9, the significant difference in this case merely reflects a difference in the relative contribution of the Oculomotor subscale to the reported sickness, not a profile difference (i.e., both studies have a (O>N>D) profile. A comparison of the individual study profiles in Figure 9 and the profile for the mean Fixed-Wing simulator (cf. Figure 6), also reveals that Study 302 is the only profile that differs from the average profile for this type of simulator. Moreover, the profile for seven of the eight studies are similar to the average Fixed-Wing simulator profile, which indicates that the average profile provides a fairly accurate reflection of the individual study profiles.

In terms of differences in the severity of sickness, the post hoc test results revealed that none of the study pairs which showed differences in the proportional subscales also had significant differences on the regular SSQ subscale scores. Moreover, none of the Fixed-Wing studies showed a significant difference on the Disorientation subscale whereas the study pairs that were significantly different, only differed on one of the subscales. In particular, only one study pair showed a difference on the Nausea subscale (Study 302 and 318) and on Oculomotor subscale, only six study pairs were significantly different. However, five of these comparisons included Study 303.

Taken together, these results indicate that within the various types of Fixed-Wing simulators, there are no quantitative differences in the SSQ profiles and only a few quantitative differences in symptom severity, which are predominately differences in Oculomotor discomfort.

Rotary-Wing Simulators

The results of the multiple comparison tests for the proportional scores indicated that there were no significant differences for any of the Fixed-Wing studies on the Oculomotor and Disorientation subscales and only one study pair differed on the Nausea subscale (Study 305 and 309). The profiles for each of the Rotary-Wing simulator studies are shown in Figure 10 below.

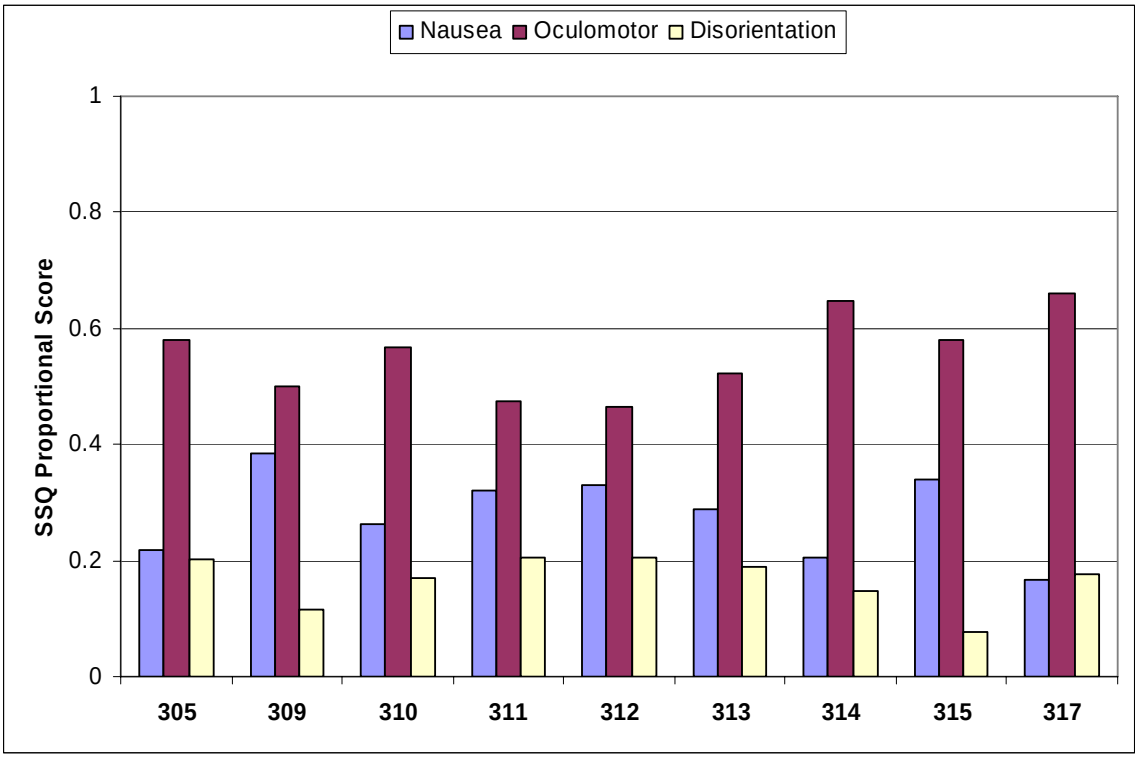

Figure 10. SSQ profiles for Rotary-Wing simulator studies.

The profiles shown in Figure 10 reflect the post hoc test results which indicate that none of the Rotary-Wing simulators have significantly different profiles (O>N>D). The results for the only significant study pair (Study 305 and 309) differed on the Nausea subscale. As with the Fixed-Wing profiles, this significant difference does not indicate a difference in profile, merely that Study 309 has a significantly higher relative contribution of Nausea symptoms in the overall reported sickness compares to Study 305. Additionally, a comparison of the individual study profiles in Figure 10 and the profile for the mean Rotary-Wing simulator (cf. Figure 6) indicates that the average profile provides an accurate reflection of the individual study profiles.

In terms of differences in the severity of sickness, the post hoc test results revealed that Study pair 305 and 309, which showed a difference in the proportional Nausea subscale, was not significantly different on any of the regular SSQ subscale scores. Moreover, none of the Rotary-Wing studies showed a significant difference on more than one of the subscales. In particular,

only two study pairs showed a difference on the Nausea subscale, three study pairs differed on Oculomotor subscale, and another two study pairs were significantly different on the Disorientation subscale.

Taken together, these results are similar to those of the Fixed-Wing studies: within the types of Rotary-Wing simulators, there are no quantitative differences in the SSQ profiles and only a few quantitative differences in symptom severity on the Nausea, Oculomotor, and Disorientation subscales.

Between System Comparisons for Individual Simulator Studies

The results of the multiple comparison tests for the proportional scores indicated that several of the between system simulator study comparisons had significant differences on at least two of the subscales. First, the profiles for the studies that differed on both the Nausea and Oculomotor subscales are presented below Figure 11.

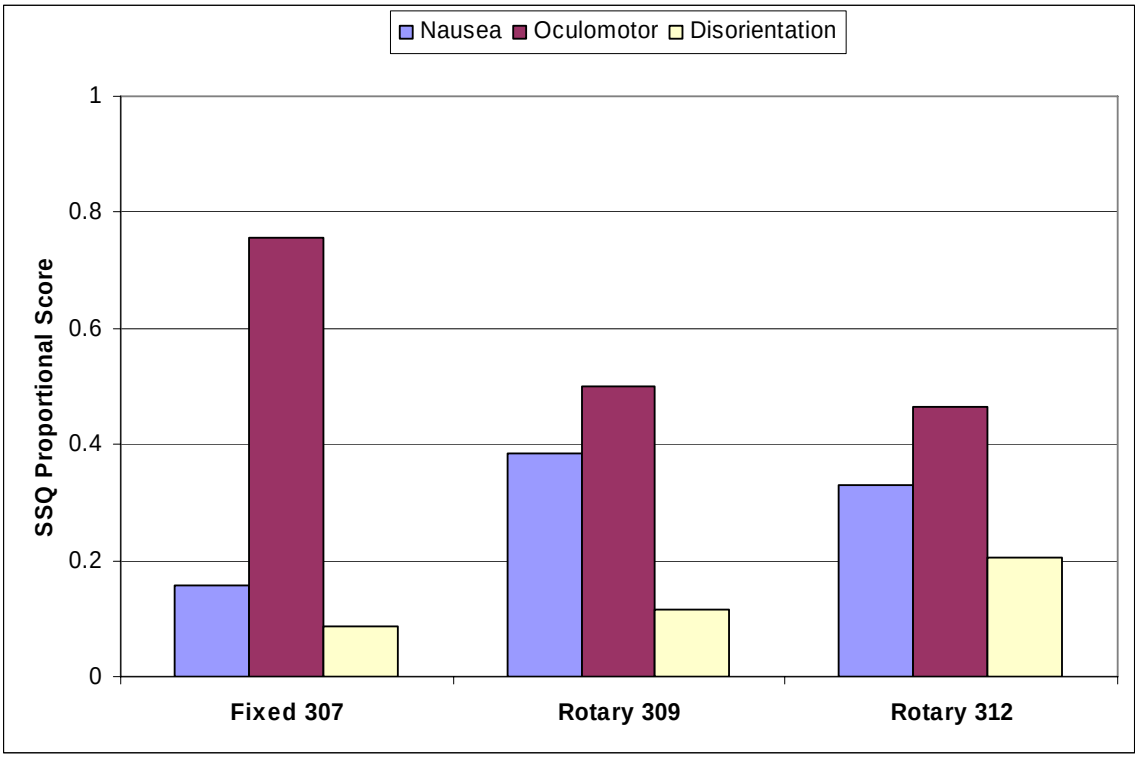

Figure 11. SSQ profiles for simulator studies with different Nausea and Oculomotor proportional scores.

The results of the post hoc analyses indicated that the Fixed-Wing simulator in Study 307 was significantly different than the Rotary-Wing simulators in Studies 309 and 312 on both the Nausea and Oculomotor subscales, which suggested that Study 307 had a different overall profile than the other two studies. However, the profiles for these studies, shown in Figure 11, all appear to have a similar profile (i.e., O>N>D). Therefore, the significant result for these two study comparisons reflect a difference in the relative contribution of Nausea and Oculomotor symptoms to the overall level of reported sickness. In particular, the Fixed-Wing simulator in Study 307 has a lower contribution of Nausea symptoms compared to the two Rotary-Wing simulators. Conversely, Study 307 has a higher relative contribution of Oculomotor symptoms than the Rotary-Wing simulators.

The post hoc test results on the regular SSQ subscales also revealed differences in the severity of sickness between Fixed- and Rotary-Wing simulators. However, only the comparison between Study 307 and 312, which showed differences in the proportional Nausea and Oculomotor subscales also had significant differences on two of the regular SSQ subscale scores. On the regular subscale comparisons, Study 307 and 312 they showed differences in sickness severity on the Nausea and Disorientation subscale. Other study comparisons also showed a significant difference on at least two of the subscales. Two of the Fixed- and Rotary-Wing study pairs (Study 304 and 312 and Study 318 and 310) differed on both the Nausea and Oculomotor subscales, whereas three of the study pairs showed differences in symptom severity on all three of the subscales (Studies 306 and 312, 318 and 312, and 316 and 312).

Differences on both the Oculomotor and Disorientation subscales were also revealed in the comparisons between the Driving simulator studies and some of the Fixed- and Rotary-Wing simulator studies (see Figure 12 below).

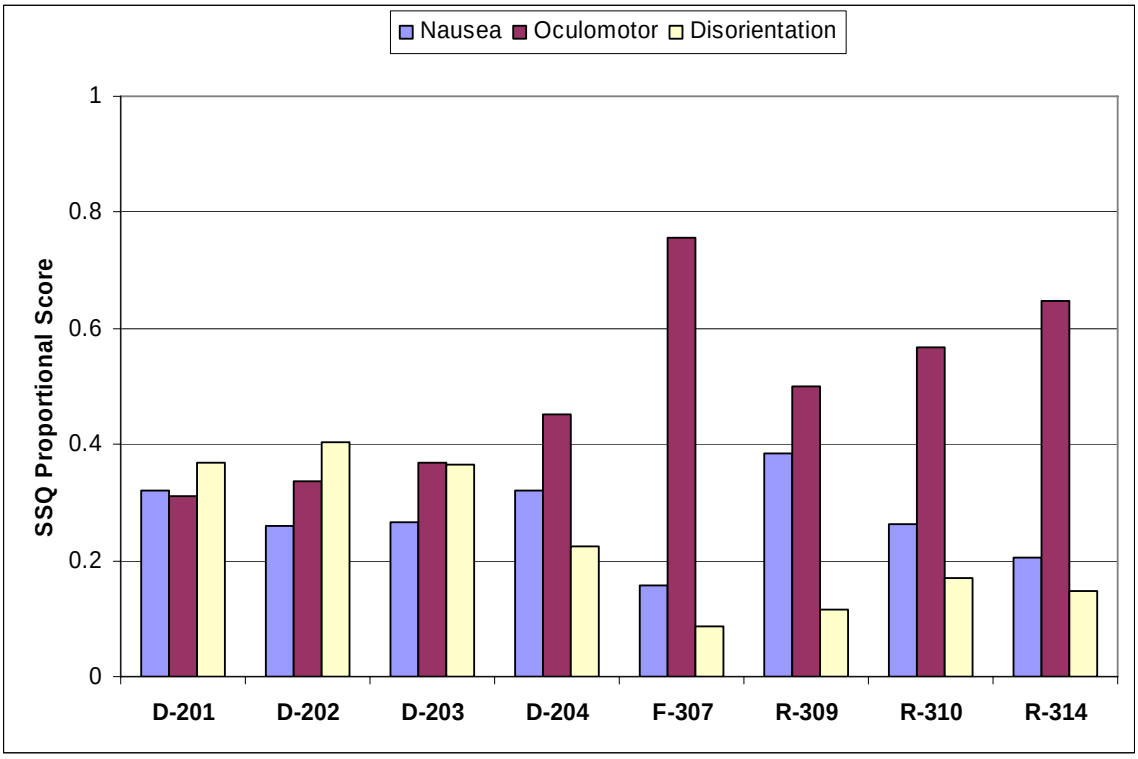

Figure 12. SSQ profiles for simulator studies with different Oculomotor and Disorientation proportional scores.

The results of the post hoc analyses indicated that the Fixed-Wing simulator in Study 307 was significantly different than the Driving simulators in Studies 201, 202, and 204 on the Nausea and Oculomotor subscales. Similarly, the Rotary-Wing simulator in Study 314 was significantly different than the Driving simulators in Studies 201, 202, and 203. Two other Rotary-Wing and Driving simulator comparisons were also significantly different on these two subscales (Studies 309 and 201 and Studies 310 and 202).

In contrast to the results for the Fixed- and Rotary-Wing comparisons, Figure 12 confirms that the significant results for these study comparisons reflect differences in profiles. Specifically, the Rotary-Wing simulator in Study 309 has a different profile (O>N>D) than the profile for the Driving simulator in Study 201 (D>N>O) and the profile for Rotary-Wing Study 310 (O>N>D) differs from the Driving Study 202 (D>O>N). Likewise, the profile for the

Rotary-Wing Study 314 (O>N>D) is different than the profiles for the Driving simulators in Studies 201 (D>N>O), 202 (D>O>N), and 203 (O>D>N). Moreover, the profile for the Fixed-Wing simulator in Study 307 (O>N>D) is different than the profiles for the Driving simulators in Studies 201 (D>N>O) and 202 (D>O>N), but Study 204 appears to have a similar profile (O>N>D). Therefore, this significant comparison (i.e., Study 307 and 204) merely reflects that Study 307 has a larger relative contribution of Oculomotor symptoms and a smaller contribution of Disorientation symptoms.

In terms of differences in the severity of sickness, the post hoc test results revealed that many of the study comparisons between the Driving simulators and the Fixed- and Rotary-Wing simulators which showed profile differences also had significant differences on at least two of the regular SSQ subscale scores. In particular, two study pairs (Study 202 and 307 and Study 202 and 310) differed on all three of the SSQ subscales whereas three of the study pairs showed a difference on the Nausea and Disorientation subscales. (i.e., Studies 201 and 307, 201 and 309, and 204 and 307). In particular, only one study pair showed a difference on the Nausea subscale (Study 302 and 318) and on Oculomotor subscale, only six study pairs were significantly different. However, five of these comparisons included Study 303.

The post hoc results also indicated that many of other study comparisons between Driving simulators and the Fixed- and Rotary-Wing simulators showed a significant difference in symptoms severity in symptom severity on at least two of the subscales. For instance, the Driving simulator in Study 202 was different than seven of the other Rotary-Wing studies and six of the other Fixed-Wing studies on all three of the SSQ subscales. Similarly, the Driving simulator in Study 201 differed from Fixed-Wing Study 304 on all three of the subscales. The Driving simulator in Study 202 also showed significant differences in symptom severity on the

Oculomotor and Disorientation subscales with the Rotary-Wing simulator in Study 312. Differences in severity on the Nausea and Disorientation subscales were found for the comparisons between Driving simulator Study 201 and two Rotary-Wing studies (Study 305 and 315) and three Fixed-Wing studies (Studies 306, 316, and 318).

Taken together, these results indicate that there are both quantitative differences in the SSQ profiles and symptom severity between the various types of simulators.

Profile and Sickness Severity Comparison for VR Studies

The results of the MANOVA for the proportional subscales scores revealed a statistically significant difference among the 16 VR studies. Similarly, the One-Way ANOVAs for the SSQ Total Severity score and the subscale scores also indicated that there were statistically significant differences among the studies. A discussion of the post hoc multiple comparison tests for all of these analyses are provided below; the within system results discussion is presented first followed by the between system results.

HMD Systems

The results of the multiple comparison tests for the proportional scores indicated that there were no significant differences for any of the HMD studies on the Nausea subscale. In contrast, two study pairs (Study 107 and 113 and Study 111 and 113) had significant differences on the Oculomotor and Disorientation subscales. Additionally, two study pairs (Study 105 and 107, Study 107 and 110) differed only on the Oculomotor subscale while Study 113 differed from four other studies (Studies 108, 109, 110, and 112) only on the Disorientation subscale. The profiles for each of the HMD studies are shown in Figure 13 below.

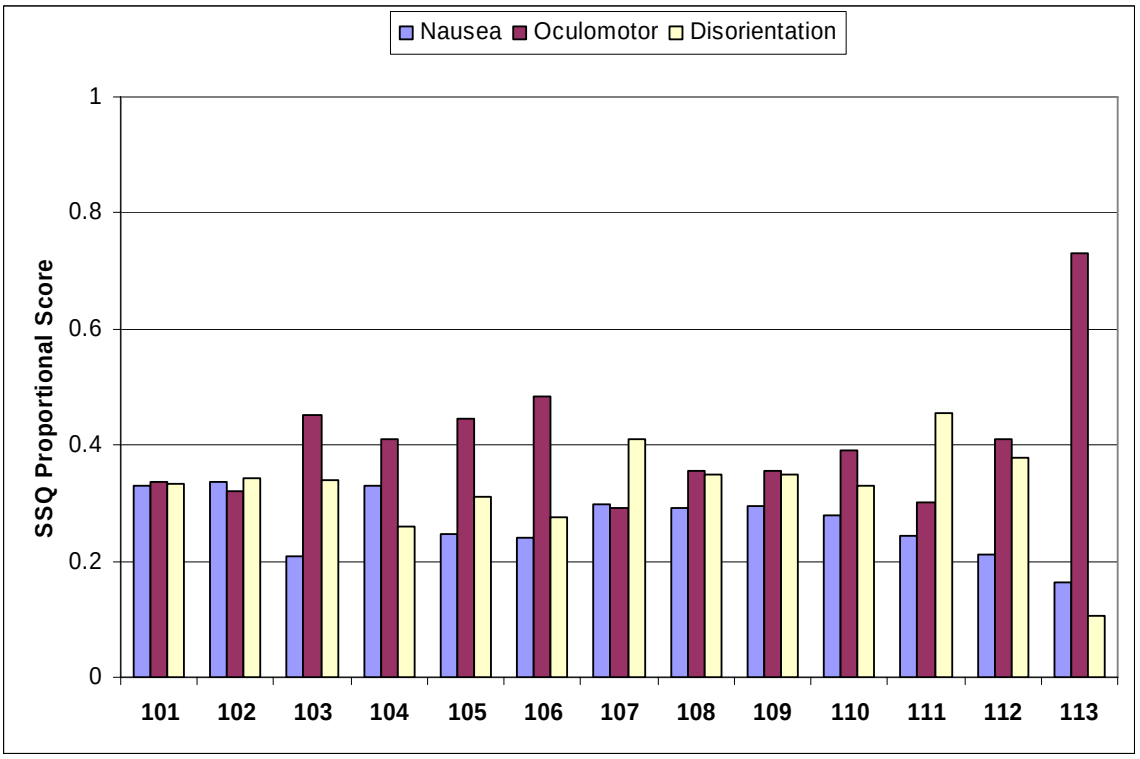

Figure 13. SSQ profiles for individual HMD studies.

The profiles shown in Figure 13 reflect the results of the post hoc tests. The significant HMD study comparisons that differed on two of the subscales (Oculomotor and Disorientation) have different profiles: Studies 107 (D>N>O) and 113 (O>N>D) and Studies 111 (D>O>N) and 113 (O>N>D). Additionally, the profiles in Figure 13 show that the study pairs which only differed on one of the subscales also have different profiles. The profiles for studies that differed on the Oculomotor subscale indicate that the difference between the studies is reflected in the position of the Oculomotor component relative to the other symptom subscales. Specifically, Study pair 105 (O>D>N) and 107 (D>N>O) and Study pair 107 (D>N>O) and 110 (O>D>N) both show the Oculomotor subscale in opposite positions within the profile. Similarly, the profile for the four studies (Study 108, 109, 110, and 112), which differed from Study 113 only

on the Disorientation subscale, also show differences between the relative position of the Disorientation component (i.e., O>D>N and O>N>D, respectively).

A comparison of the individual study profiles in Figure 13 with the mean HMD profile shown previously in Figure 7 reveal that eight of the 13 (62%) individual HMD studies have a similar profile (i.e., O>D>N). Therefore, although the results revealed that there are several individual studies which have different profiles, the average HMD profile provides a fairly accurate reflection of the individual study profiles.

As with the differences in profiles between Study 107 and 113, there were also significant differences for this study pair on two of the SSQ subscale severity scores (Nausea and Disorientation), in which the means for Study 107 were greater than Study 113. Moreover, Study pair 110 and 113, which showed a difference in profile as a result of the significant difference between their mean proportional Disorientation scores, also had significant differences in symptom severity on all three of the regular SSQ subscales. The results for this study pair showed that the mean score for Study 110 was significantly greater than Study 113 on all of the subscales (N, O, and D).

Several other study pairs showed significant differences in sickness severity on all three of the SSQ subscales, but did not show differences on the proportional scores. In particular, Study 111 had a significantly lower mean on all three of the subscales compared to Studies 105, 109, and 110. The mean for all three of the SSQ subscales was also significantly lower than for Study 112 compared to five of the other HMD studies (Study 105, 107, 108, 109, and 110). Likewise, Study 113 had a significantly lower mean than Study 105 on all of the subscales.

Taken together, these results indicate that there are quantitative differences in the SSQ profiles and symptom severity within the various types of HMD systems.

BOOM and CAVE Systems

The results of the multiple comparison tests for the proportional scores indicated that there were no statistically significant differences between the two BOOM studies on any of the proportional subscales. Additionally, since there was only one CAVE system study represented in the database, no within system comparisons were possible. The profiles for the two BOOM studies and the CAVE study are shown in Figure 14 below.

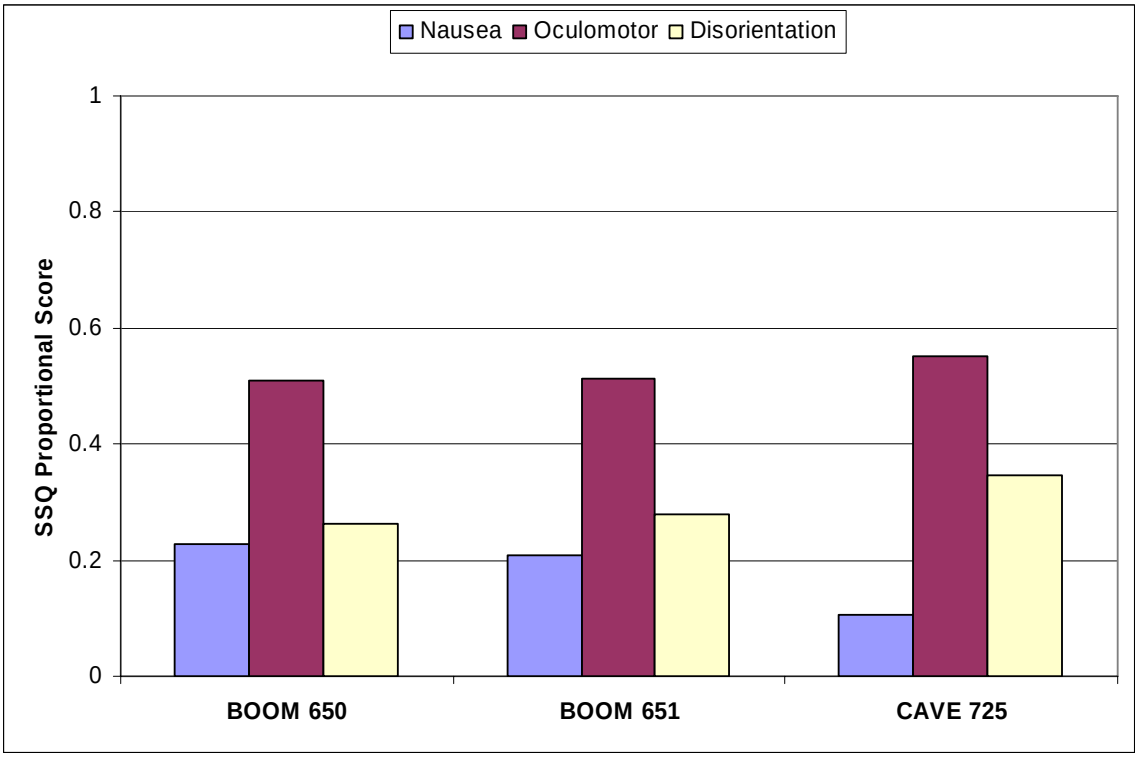

Figure 14. SSQ profiles for individual BOOM and CAVE studies.

Between System Comparisons for the Individual VR Studies

The results of the multiple comparison tests for the proportional scores revealed that none of the between system VR study comparisons had significant differences on more than one of the subscales. However, the analyses did indicate that one of the HMD studies (Study 107) had a

significantly lower mean proportional score on the Oculomotor subscale than one of the BOOM studies (Study 651), which resulted in a different profile between the two systems. As shown in Figure 15, the profile for the HMD in Study 107 was D>N>O whereas the BOOM in Study 651 had a O>D>N profile.

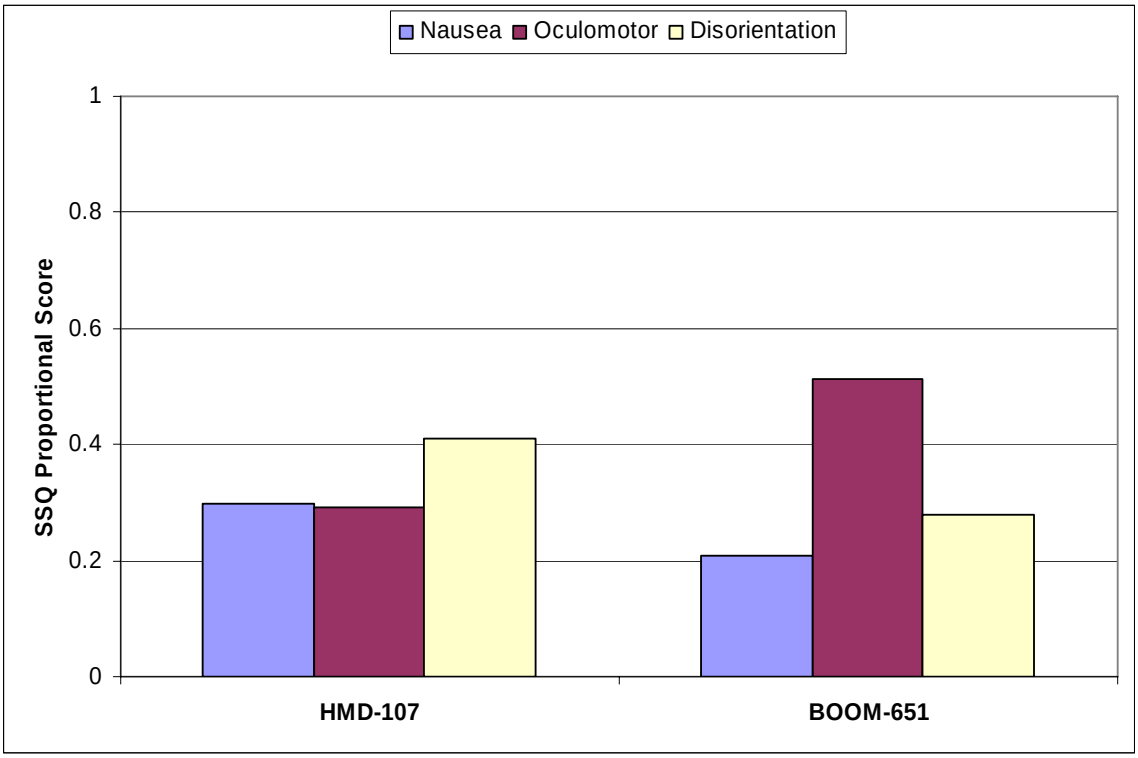

Figure 15. SSQ profiles comparing an HMD and BOOM study.

As with the difference in profiles between the HMD in Study 107 and the BOOM in Study 651, there were also significant differences for this study pair on two of the SSQ subscale severity scores (Nausea and Disorientation), in which the means for the BOOM were lower than the HMD system. Several other HMD-BOOM study pairs showed significant differences in sickness severity on the Nausea and Disorientation SSQ subscales, but did not show differences on the proportional scores. In particular, the BOOM in Study 651 had a significantly lower mean on both of these subscales compared to the HMDs in Studies 108, 109, and 110.

The results of the multiple comparison also revealed that the CAVE system had a significantly lower mean proportional score on the Nausea subscale compared to the mean for five of the HMD studies (Study 101, 107, 108, 109, and 110). However, as shown in Figure 16 below, only the HMD in Study 107 had a different profile (D>N>O) than the profile for the CAVE system (O>D>N). The significant differences between the CAVE system and the other HMD studies merely indicated that the relative contribution of the Nausea symptoms to the overall level of sickness reported was lower for the CAVE system.

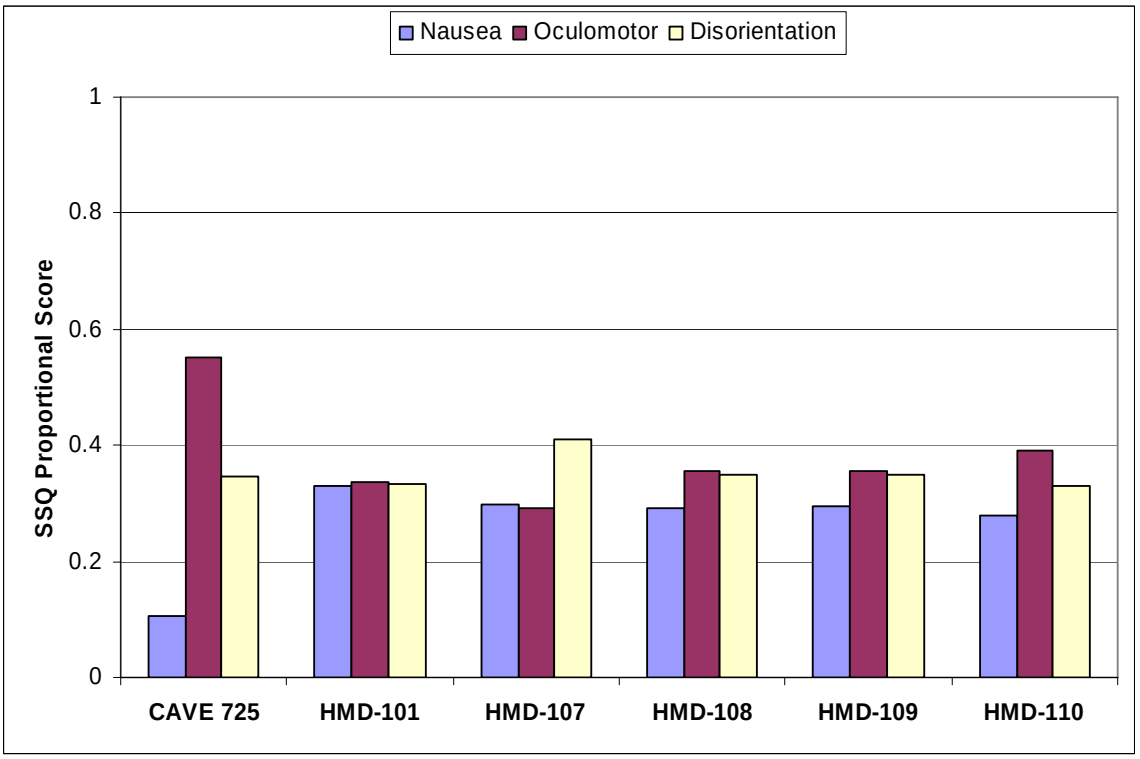

Figure 16. SSQ profiles comparing the CAVE study to five HMD studies.

As with the difference in profiles between Study 107 and the CAVE study, there were also significant differences for this study pair on all three of the SSQ subscale severity scores, in which the means for Study 107 were greater than the CAVE system. The comparisons between the CAVE system and the HMDs in Studies 108, 109, and 110, which showed a difference in the

proportional Nausea subscale score, had significant differences in sickness severity on all three of the SSQ subscales. Additionally, although no significant differences were found on any of the proportional subscales between the HMD in Study 105 and the CAVE, significant differences in sickness severity were also found on all three of the subscales.

Several other HMD-CAVE study pairs showed significant differences in sickness severity on two of the SSQ subscales, but did not show differences on the proportional scores. In particular, Studies 101, 102, and 106 had a significantly greater mean on the Nausea and Oculomotor subscales compared to the CAVE system. Likewise, Study 113 had a significantly lower mean than Study 105 on all of the subscales. Finally, the post hoc results revealed a difference in sickness severity between the BOOM in Study 651 and the CAVE study on the Nausea and Oculomotor subscales. In these comparisons, the severity of Nausea symptoms was greater in the BOOM system, but the Oculomotor symptoms were greater in the CAVE system.

Taken together, these results also indicate that there are quantitative differences in the SSQ profiles and symptom severity within the various types of VR systems.

Conclusions

The results of the research showed statistically significant differences in the SSQ profiles and the overall severity of sickness between simulator and VR systems, which provide evidence that simulator sickness and VR sickness represent distinct forms of motion sickness. Accordingly, these results support the use of two separate terms (i.e., simulator sickness and cybersickness) to refer to the negative side-effects of exposure to these systems.

Analyses on three types of simulators (i.e., Fixed- and Rotary-Wing flight simulators and Driving simulators) also found significant differences in the sickness profiles as well as the

overall severity of sickness within different types of simulator systems. Additional analyses on the data from the individual simulator studies confirmed the differences in SSQ profiles between the various types of simulator systems and found differences in symptom severity between the three types of simulators. While the results also revealed quantitative differences in the SSQ profiles and symptom severity within the various types of Driving simulators, no differences in the SSQ profiles were found within the various types of Fixed-Wing and Rotary-Wing simulators.

A review of the significant study comparisons for each of the proportional subscales revealed some commonalities among the equipment features that differed between the simulator studies. In all of the significant comparisons on the proportional Nausea subscale, the study simulator with a greater mean score had a larger field-of-view. However, the significant comparisons on the proportional Oculomotor subscale showed the opposite effect for field-of-view. Specifically, in 79% of the significant study pairs, the study with a greater mean proportional Oculomotor score had a smaller field-of-view. Finally, the equipment features that differed for the significant study comparisons on the proportional Disorientation subscale indicated that in 84% of the study pairs, larger mean scores were noted in the simulator that did not have a motion base. Additionally, 63% of the significant study pairs had differences in the type of display. A review of these study pairs revealed that in all of the Projection Screen-Dome display differences, the study with the projection screen had a greater mean proportional Disorientation score. Similarly, in 88% of the Projection Screen-CRT display differences, the study with the projection screen also had a greater mean score. Finally, in 67% of the Dome-CRT display differences, the study with the Dome display had a greater mean score. Although

many of the significant study comparisons also differed in the display's field-of-view, half of the studies with a greater mean score had a larger field-of-view and half had a smaller field-of view.

Analyses on three types of VR systems (i.e., HMD, BOOM, and CAVE) revealed that BOOM and CAVE systems have similar sickness profiles, which are different than the HMD system profile. Moreover, the results showed that the overall severity of sickness is greater in HMD systems than in BOOM and CAVE systems. Analyses on the data from the individual VR studies confirmed the differences in SSQ profiles between HMD systems and BOOM and CAVE systems. The results also showed significant differences in SSQ profiles and symptom severity within the various types of HMD systems. However, no differences in SSQ profiles or symptom severity were found within the BOOM studies.

A review of the significant study comparisons for each of the proportional subscales revealed some commonalities among the equipment features that differed between the VR studies. In all of the significant comparisons on the proportional Nausea subscale, the study with the greater mean score had a smaller field-of-view, slower speed of the head tracker, and the weight of the display was larger. Comparisons of the equipment features for the significant results on the proportional Oculomotor subscale showed that in both of the study pairs where differences in field-of-view were noted, the study with the greater mean score had a larger field-of-view. Additionally, in two of the three study pairs where differences in the head tracker speed were noted, the study with a greater mean proportional Oculomotor score had a faster head tracker. Finally, the equipment feature differences for the significant study comparisons on the proportional Disorientation subscale indicated that in 67% of the study pairs, a faster head tracker speed was also noted in the study with the greater mean score. Although two study pairs also differed in weight, one study pair showed that the greater mean proportional Disorientation

score occurred in the study where the weight of the display was larger whereas the other study pair showed the opposite effect. A similar problem was also noted for the two study pairs where the latency of the head tracker differed.

At this time, the relationships between the engineering characteristics of VE systems and specific types of sickness symptoms that were identified in the preceding paragraphs are only speculative due to the nature of the research. However, they do provide testable hypotheses regarding the equipment features that can be evaluated in future research applications in order to ultimately identify which design features are best suited to minimize particular types of symptoms.

Unlike previous VE studies in which the results of the SSQ subscales were reported, this research used a new method, proportional subscale scores, to evaluate differences in the symptoms profiles among each of the individual studies. The proportional scores "normalized" the subscale scores relative to the sum of the three subscale scores to reflect only the relative contribution of the subscale scores. Accordingly, the transformation of the subscale scores into the proportional subscale scores provides a means to identify which of the subscales have similar profiles and also which of the subscales tend to dominate different types of systems or equipment features regardless of total severity. While the original SSQ subscale scores were still used to evaluate differences in the severity of sickness, the proportional scores were used to create the profiles for each VE study and to evaluate differences in profiles between VE systems and individual VE studies.

Overall, the expected relationship between symptom profiles and the type of VE system that produced them was borne out by the analyses. The results showed quantitative differences in the SSQ profiles and severity of sickness both within and between the different types of VE

systems. The results of the research also revealed statistically significant differences in the SSQ profiles and the overall severity of sickness between simulator and VR systems, which provide evidence that simulator sickness and VR sickness represent distinct forms of motion sickness. Accordingly, these results support the use of two separate terms (i.e., simulator sickness and cybersickness) to refer to the negative side-effects of exposure to these systems.

Kennedy, Drexler, Stanney, and Harm (1997) suggested that SSQ profile differences (e.g., excessive visual disturbance) may signal differences in specific equipment design features that differentially affect the severity and types of symptoms reported. Accordingly, another goal of the research was to determine the relationship between different engineering features and the SSQ symptom subscales for different types of VE systems. Although potential system variables that may influence sickness were identified for the systems which had significant profile differences, it was not possible to establish definitive relationships in this phase of the research. Even after "cleaning" the data, the final database used in the analyses was exceptionally large (2,100 individuals) and represented a variety of equipment configurations for both types of VE systems (simulators and VRs). However, at this point one cannot rule out the possibility that additional data could have provided more definitive conclusions about the affect of various equipment features on different types of sickness. It is also possible that data from a control group could have assisted in identifying more conclusions about the relationship between equipment features and SSQ profiles. Relatedly, several comparisons of the equipment features for study pairs that were identified as significantly different revealed that there were no differences in the equipment between the two studies. This finding indicates that some other factor in the studies may have been responsible for the difference in reported sickness. Although the source of the difference between these studies is unknown, possible factors include exposure

duration or characteristics of the participants (e.g., differences in susceptibility to motion sickness, different ages, different levels of experience in provocative motion environments, etc.). Another possible reason for this finding relates to the statistical analyses. Specifically, because some of the multiple comparison tests had fewer significant differences than might be expected with the alpha level that was used in the analyses, some of the significant differences that were identified may have been spurious.

Limitations of the Research

In a typical research study, the principle investigator manipulates the independent variable under investigation and controls the influence of extraneous variables either directly, in the study design, or indirectly through randomization. Therefore, any differences that are revealed in the research results can be attributed to the independent variable. In this research, however, control over extraneous variables that could influence the sickness symptoms reported in the individual studies was obviously not possible and as a result, unambiguous interpretation of any differences found between VE systems was not possible. Accordingly, any conclusions regarding the equipment features that were responsible for, or at least contributed to, significant differences in reported symptoms between the VE studies could only be speculative.

An unanticipated problem that was encountered during this research was the lack of willingness by many researchers to contribute their SSQ data to this study. Obviously, in a study of this nature (i.e., evaluating profiles from various VE systems), a substantial amount of data is required; much more than a single researcher could accumulate in any reasonable number of years. Historically, researchers readily agreed to provide any data collected with the SSQ as a proviso for permission to use the questionnaire and any assistance with scoring questions.

During this research, a majority of these researchers were contacted and asked to contribute their SSQ data (along with a citation for their study in order to provide an appropriate attribution for their data) for inclusion in this research. Several scientists were very amenable to the request and forwarded their SSQ data including an offer to provide any additional information that was needed for the project (cf. Appendix D and E). In stark contrast, there were a few scientists that outright refused to provide their data while others took a more passive approach in their refusal and simply ignored repeated requests to include their SSQ data in this research. These negative responses were not only disappointing, they were also rather surprising since the research community touts a cooperative atmosphere in which scientists share their research (e.g., journal articles, scientific conferences) in order to advance the research in their respective fields. Moreover, several of the scientists that refused to contribute their data were long-time colleagues of Dr. Robert Kennedy, one of the developers of the SSQ and the person that was actually requesting the data from them. The scientists and practitioners in the simulator and VR community all concur that the sickness associated with exposure to simulated environments is a considerable problem that impedes advancement of the technology as well as existing and future VE applications. However, based on the number of negative responses that were received during this research, simulator and cybersickness will remain an unresolved problem until more members of the research community actually adopt the cooperative attitude that they proclaim to possess.

Future Research

There are many opportunities for extending this research to further contribute toward understanding the differential effects of various equipment features on sickness outcomes in

order to facilitate effective management of VE-induced sickness (i.e., minimize side effects). As stated previously, Kennedy, Lanham, Drexler, Massey, and Lilienthal (1997) suggested that the first technical step toward improving VE systems so that they do not induce sickness is to quantify, as accurately as possible, the problem(s) that are experienced by the people who use them. This research provided that first step by identifying potential system design characteristics that influence the symptoms experienced by VE system users. However, additional empirical research is needed to test the system design hypotheses. In particular, psychophysical studies are needed to evaluate the relationship between different combinations of equipment features and the specific types of sickness symptoms that are produced by exposure to the system. For example, the results from the VR studies showed that higher proportional Nausea scores occurred in systems with smaller fields-of-view, but proportional Oculomotor scores showed an opposite relationship with field-of-view (i.e., lower scores were found in systems with smaller fields-of-view). Therefore, a future study could identify the SSQ profile and severity of sickness associated with a particular VR system and then after modifications are made to the field-of-view, examine the data from the modified system to determine the effects, if any, on the profile and sickness severity.

The results of the current research also revealed quantitative differences in sickness profiles and severity between simulator and VR systems (i.e., simulator sickness and cybersickness). While this difference may be due to differences in the equipment features of the two systems, it has been previously suggested that the differences could also be due to a population difference. Specifically, Kennedy, Drexler et al. (2003) noted that the sickness data for flight simulators were collected from military pilots whereas the VR participants were primarily college students. The authors indicated that due to the nature of their occupation,

military pilots are generally self-selected as being more immune to motion sickness and they are more likely to underreport sickness symptoms compared to college students, both of which could affect SSQ scores. Consequently, until studies are conducted which address this potential sample bias, it cannot be ruled out as a contributing factor in the sickness differences between the two types of VE systems. An approach to determine whether the differences between simulator and cybersickness are truly different or merely an artifact of the differences between military personnel and college students would be to collect SSQ data from military pilots exposed to VR systems and then compare their scores to those collected from college students.

Relatedly, Lane and Kennedy (1988) originally developed the SSQ because differences between traditional motion sickness and simulator sickness (e.g., less severe symptoms) suggested that the Motion Sickness Questionnaire (MSQ) was not an ideal measure of simulator sickness. They also noted that some of the symptoms which were valid in the MSQ scoring method were not appropriate for measuring simulator sickness because they were rarely reported in simulator exposures. If additional research indicates that VR sickness is sufficiently different from simulator sickness, the use of a separate measurement instrument may be warranted. Therefore, future research could include a factor analysis of the SSQ data collected after VR exposure in order to create a modified version of the SSQ that is specifically designed to quantify sickness related to VR exposure.

Other opportunities for extending this research to further contribute toward an understanding of deleterious side effects of VE exposure relate to situations where multiple exposures to a particular VE stimulus are required (e.g., training applications). Previous research has shown that repeated exposures to flight simulators generally reduce the severity of sickness in subsequent exposures (Kennedy, Berbaum, Dunlap, & Smith, 1995; Kennedy, Hettinger, &

Lilienthal, 1990). However, whether adaptation is affected by the characteristics of the equipment is still an open question. Examples of research questions in this area include: is adaptation affected by the size of the field-of view (e.g., narrow vs. wide FOV); if adaptation is affected by a particular aspect of the equipment, does it facilitate or hinder the adaptation process; and if there is an adaptation effect, is it specific to a particular type of VE system (i.e., only simulators or only VRs). Results from investigations of this nature would provide important information for the design of usage schedules for VE systems (i.e., the amount of time between subsequent exposures to the same system).

Finally, research is needed to address an unresolved methodological problem that exists in studies which evaluate simulator sickness and cybersickness. In any type of research study, an investigator is ethically required to allow participants to withdraw from their study at any time and for any reason. For sickness research, the question that researchers then face is how to handle the data for the individuals that withdrew from their study. Many researchers simply remove the data from these participants and either analyze the data for a smaller number of study participants or they run additional participants to replace the missing data. However, participants that remain in the study (i.e., don't drop out due to sickness) are essentially self-selected as not susceptible, or less susceptible to sickness. Therefore, only analyzing the data for these participants would not only fail to capture the effects on the general population of potential users, it could also show that the incidence or severity of sickness associated with exposure to the particular device is lower than what would be found if the "drop-out" data were included in the analyses. Moreover, the number of participants that withdraw from a study due to sickness can provide important information about the source of a sickness problem. In particular, many participant withdrawals suggest a problem with the VE system itself whereas only a few

participant withdrawals suggest differences in the system user (e.g., the participants withdrew because they were more susceptible to sickness). Accordingly, excluding the data of participants that withdraw from a study could lead to erroneous conclusions regarding the sickness associated with exposure to the system.

APPENDIX A:
COMPUTATION OF THE SSQ SCORES

	SSQ Subscales		
SSQ Symptom[1]	**Nausea (N)**	**Oculomotor (O)**	**Disorientation (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[2]	[1]	[2]	[3]

Score

N = [1] x 9.54

O = [2] x 7.58

D = [3] x 13.92

TS[3] = ([1] + [2] + [3]) x 3.74

[1] Scored 0, 1, 2, 3; [2] Sum obtained by adding symptom scores; [3] Total Severity score

APPENDIX B:
SIMULATOR STUDIES

Study	System Type	Display Type	Study Name	Study Location	Year	Aircraft/ Envir	Avg Latency (ms)	Image Generator	Res-H	Res-V	FOV-H	FOV-V	Motion Base	Motion DOF
201	Driving	Projection Screen	UCF Driver Training Simlator	Orlando, FL	1998	Dodge Aries 4-door cab	50	Silicon Graphics Inc (SGI) Onyx Reality Engine 2	1920	480	160	45	No	.
202	Driving	CRT	Ford Driving Simulator (FDS)	Ford Motor Company	2002	Full-size vehicle	80	Evans & Sutherland ESIG 2000	3150	900	140	40	Fixed	.
203	Driving	Dome	VIRtual Test Track Exper. (VIRTTEX)	Ford Motor Company	2002	Full-size vehicle	70	Evans & Sutherland ESIG 2000	7200	1600	180	40	Yes	6
204	Driving	Projection Screen	Highway Driving Sim (HDS)	Fed. Hwy. Admin.	2003	4-Door Saturn sedan cab	.	SGI Onyx2/Infinite Reality 2 (IR2)	1920	1200	88	.	Yes	3
302	Fixed	Dome	2E7	NAS LeMoore, CA	1984	F/A-18 / WTT	.	Digital CGI	.	.	360	145	No	.
303	Fixed	CRT	2F110	NAS Miramar, CA	1984	E-2C (Hawkeye) / OFT	.	Digital CGI/Hybrid CRT	.	.	139	35	Yes	6
304	Fixed	Dome	2F112	NAS Miramar, CA	1984	F-14A (Tomcat) / WST	.	TV camera carrier model; Point light	.	.	360	150	No	.
305	Rotary	CRT	2F117	MCAS New River, NC	1984	CH-46E (Sea Night) / WST	200	Evans & Sutherland CT-5	.	.	175	50	Yes	6
306	Fixed	CRT	2F132	NAS LeMoore, CA	1984	F/A-18 (Hornet) / OFT	.	Calligraphic CGI	.	.	48	32	No	.
307	Fixed	CRT	2F87F	NAS Brunswick, GA	1984	P3-C (Orion) / WST	150	McDonnel Douglas Vital IV CIG	.	.	48	36	Yes	6
308	Fixed	CRT	2F87F	NAS Jacksonville, FL	1986	P3-C (Orion) / WST	.	TV camera/ Model Board Projection	.	.	48	36	Yes	6

309	Rotary	CRT	2F121	MCAS New River, NC	1984	CH-53D (Stallion) / OFT	.	Evans & Sutherland CT-5	.	.	200	50	Yes	6
310	Rotary	CRT	2F120	MCAS Tustin, CA	1984	CH-53E (Super Stallion) / OFT	177	Evans & Sutherland CT-5	.	.	200	50	Yes	6
311	Rotary	CRT	2F120	MCAS New River, NC	1985 1990	CH-53E (Super Stallion) / OFT	177	Evans & Sutherland CT-5A	.	.	200	50	Yes	6
312	Rotary	CRT	2F64C	NAS Jacksonville, FL	1985	SH-3 (Sea King) / WST	215	McDonnel Douglas Vital IV CIG	.	.	130	30	Yes	6
313	Rotary	CRT	2B33	Ft Rucker, AL	1989	AH-1S (Cobra) / FWS	.	Digital image generator/ collimating mirrors	.	.	48	36	Yes	6
314	Rotary	CRT	2B31	Ft. Campbell, KY	1989	CH-47D (Chinook)	.		.	.	48	36	Yes	6
315	Rotary	CRT	2F120	MCAS Tustin, CA	1991 1992	CH-53E (Super Stallion)/ OFT	177	Evans & Sutherland CT-5	.	.	200	50	Yes	6
316	Fixed	Dome	2F143	NAS Whidbey Island, WA	1991	EA-6B (Prowler) / OFT	.	Evans & Sutherland ESIG500 SPX	.	.	180	45	Yes	6
317	Rotary	CRT	2F117A	MCAS Tustin, CA	1992	CH-46E (Sea Knight)/ OFT	.	Evans & Sutherland CT-5A	.	.	200	50	Yes	6
318	Fixed	Dome	2F143	NAS Whidbey Island, WA	1992	EA-6B (Prowler) / OFT	.	Evans & Sutherland ESIG500 SPX	.	.	180	45	Yes	6

APPENDIX C:
VIRTUAL REALITY STUDIES

Study	Type	Location	Year	Model	Res-H	Res-V	FOV-H	FOV-V	FOV-D	Screen Size	Screen Type	HMDWt-oz	IPD Adjust	Head Tracker	Tracker Speed	Tracker Latency
101	HMD	UCF, Murray State Univ	1996	i*glasses! by Virtual i*O	640	480	24	18	30	0.7	LCD	8.5	No	Virtual i*O	40	.
102	HMD	Univ of Idaho	1996	VictorMax Cybermax 180	789	230	53	35	63.5	0.7	LCD	20		Yes	.	.
103	HMD	Univ of Houston	1996	Virtual Research VR-4	742	230	48	36	60	2.7	LCD	33	Yes	Polhemus 3-space fasttrack	120	4
104	HMD	UCF	1996	Kaiser Electro-Optics VIM 500HRpv	640	480	40	30	50	1.5	LCD	24.5	No	Virtual i*O	40	.
105	HMD	Orlando	1994 1995	Virtual Research Flight Helmet	360	240	50	41	64.7	.	LCD	.	No	Polhemus Isotrak	60	20
106	HMD	Orlando	1994 1995	Virtual Research Flight Helmet	360	240	50	41	64.7	.	LCD	.	No	No	.	.
107	HMD	UCF	2004	Virtual Research VR-6	640	480	48	36	60	1.3	LCD	29	Yes	Virtual i*O	40	.
108	HMD	UCF	2004	Virtual Research VR-6	640	480	48	36	60	1.3	LCD	29	Yes	Virtual i*O	40	.
109	HMD	UCF	2004	Virtual Research VR-6	640	480	48	36	60	1.3	LCD	29	Yes	Virtual i*O	40	.
110	HMD	UCF	2004	Virtual Research VR-6	640	480	48	36	60	1.3	LCD	29	Yes	Virtual i*O	40	.
111	HMD	UNC-Chapel Hill	2002	Virtual Research VR-8	640	480	48	36	60	1.3	LCD	34	Yes	3rd Tech Hi Ball 3000	160	1
112	HMD	UNC-Chapel Hill	2002	Virtual Research VR-8	640	480	48	36	60	1.3	LCD	34	Yes	3rd Tech Hi Ball 3000	160	1

113	HMD	UCF	2005	Virtual Research VR-6	640	480	48	36	60	1.3	LCD	29	Yes	Flock of Birds	144	10
650	BOOM	Orlando	1995 1998	Fakespace BOOM2C	1280	492	140	90	166	.	CRT	.		Yes	60	200
651	BOOM	Orlando	1998	Fakespace BOOM2C	1280	1024	140	90	166	.	CRT	.		No	.	.
725	CAVE	UNC-Chapel Hill	2001	CrystalEyes shutter glasses	.	.	180	120	.	.	Shutter	3.3	No	Intersense IS-900	180	4

APPENDIX D:
SIMULATOR AND VIRTUAL REALITY STUDY REFERENCES

Simulator Studies

Study Number	Reference
201	Jones, S. A. (1998). *Effects of restraint on vection and simulator sickness.* Unpublished doctoral dissertation, University of Central Florida, Orlando.
202 - 203	Curry, R., Artz, B., Cathey, L., Grant, P., & Greenberg, J. (2002). Kennedy SSQ results: Fixed- vs. motion-base Ford simulators. *Proceedings of the Driving Simulation Conference "DSC2002"* (pp. 289-299).
204	Hoffman, R. B., Molino, J. A., & Inman, V. W. (2003). Driving simulator sickness management at Turner-Fairbank Highway Research Center. *Proceedings of the Driving Simulation Conference "DSC2003"*. Davis, G. W., & Inman, V. W. (2004). Driver responses to an infrastructure based intersection collision warning system. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 1122-1125). Santa Monica, CA: Human Factors & Ergonomics Society.
302 - 307	Gower, D. W., Lilienthal, M. G., Kennedy, R. S., & Fowlkes, J. E. (1988). Simulator sickness in U.S. Army and Navy fixed- and rotary-wing flight simulators. *AGARD Conference Proceedings No. 433: Motion Cues in Flight Simulation and Simulator Induced Sickness* (pp. 8.1-8.20). Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development. Kennedy, R. S., Lilienthal, M. G., Berbaum, K. S., Baltzley, D. R., & McCauley, M. E. (1987, September). *Simulator sickness in 10 U.S. Navy flight simulators* (NAVTRASYSCEN TR-87-008). Orlando, FL: Naval Training Systems Center. Kennedy, R. S., Lilienthal, M. G., Berbaum, K. S., Baltzley, D. R., & McCauley, M. E. (1989). Simulator sickness in U.S. Navy flight simulators. *Aviation, Space, and Environmental Medicine, 60*, 10-16.
308	Kennedy, R. S., Lilienthal, M. G., Berbaum, K. S., Wilkes, R. L., & Fowlkes, J. E. (1987, March). *Simulator sickness from very low frequency vibration in ground-based flight trainers.* Report prepared for Martin Marietta Energy Systems, Inc. for the U.S. Department of Energy Contract DE-AC05-840R21400. Orlando, FL: Essex Corporation. Van Hoy, B. W., Allgood, G. O., Lilienthal, M. G., Kennedy, R. S., & Hooper, J. M. (1987). Inertial and control systems measurements of two motion-based flight simulators for evaluation of the incidence of simulator sickness. *Proceedings of the IMAGE IV Conference* (pp. 265-273). Phoenix, AZ: Image Society Incorporated.
309	Kennedy, R. S., & Smith, M. G. (1990, September). *Simulator sickness in two Marine Corps helicopter trainers: Influence of configuration, usage, and pilot history* (Final Report, Contract DAAL03-86-D-0001). Research Triangle Park, NC: Battelle, Inc.

310	Kennedy, R. S., & Smith, M. G. (1990, September). *Simulator sickness in two Marine Corps helicopter trainers: Influence of configuration, usage, and pilot history* (Final Report, Contract DAAL03-86-D-0001). Research Triangle Park, NC: Battelle, Inc. Lilienthal, M. G., Kennedy, R. S., & Hooper, J. (1987, November). *Vision motion-induced sickness in Navy flight simulators: Guidelines*. Paper presented at the 9th Interservice/Industry Training Systems Conference, Washington, D.C. Lilienthal, G. G., & Merkle, P. J. (1986). Simulator sickness in flight simulators: A case study. *Proceedings of the 65th Annual Meeting of the Transportation Research Board: Transportation Research Record 1059* (pp. 81-86).
311	Kennedy, R. S., & Smith, M. G. (1990, September). *Simulator sickness in two Marine Corps helicopter trainers: Influence of configuration, usage, and pilot history* (Final Report, Contract DAAL03-86-D-0001). Research Triangle Park, NC: Battelle, Inc.
312	Kennedy, R. S., Lilienthal, M. G., Berbaum, K. S., Wilkes, R. L., & Fowlkes, J. E. (1987, March). *Simulator sickness from very low frequency vibration in ground-based flight trainers*. Report prepared for Martin Marietta Energy Systems, Inc. for the U.S. Department of Energy Contract DE-AC05-840R21400. Orlando, FL: Essex Corporation. Lilienthal, M. G., Kennedy, R. S., & Hooper, J. (1987, November). *Vision motion-induced sickness in Navy flight simulators: Guidelines*. Paper presented at the 9th Interservice/Industry Training Systems Conference, Washington, D.C. Van Hoy, B. W., Allgood, G. O., Lilienthal, M. G., Kennedy, R. S., & Hooper, J. M. (1987). Inertial and control systems measurements of two motion-based flight simulators for evaluation of the incidence of simulator sickness. *Proceedings of the IMAGE IV Conference* (pp. 265-273). Phoenix, AZ: Image Society Incorporated.
313	Gower, D. W., Jr., & Fowlkes, J. (1989, September). *Simulator sickness in the AH-1S (Cobra) flight simulator* (Final Report, USAARL Report No. 89-20). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.
314	Gower, D. W., Jr., Fowlkes, J., & Baltzley, D. R. (1989, September). *Simulator sickness in CH-47 (Chinook) flight simulator* (Final Report, USAARL Report No. 89-28). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.
315	Kennedy, R. S., Jones, S. A., & Smith, M. G. (1992). *The effects of platform motion on the incidence of simulator sickness in two rotary-wing flight trainers* (Final Report No. NTSC TR92-017). Orlando, FL: Naval Training Systems Center. Kennedy, R. S., Lane, N. E., & Braun, C. C. (1993). *A longitudinal field investigation of simulator sickness in the CH-46E/CH-53E flight trainers: Measurement issues related to usage of Image equipment and other variables* (Technical Report, Contract No. N61339-93-C-0031 with Naval Air Warfare Center). Orlando, FL: Naval Training Systems Center.
316	Kennedy, R. S. (1993). *Investigation and analysis of simulator sickness* (Final Report, Contract No. EERS-VN-C-92-025). Vienna, VA: EER Systems Corporation.

	Kennedy, R. S., Berbaum, K. S., & Lilienthal, M. G. (1997). Disorientation and postural ataxia following flight simulation. *Aviation, Space, and Environmental Medicine, 68*(1), 13-17.
317	Kennedy, R. S., Lane, N. E., & Braun, C. C. (1993). *A longitudinal field investigation of simulator sickness in the CH-46E/CH-53E flight trainers: Measurement issues related to usage of Image equipment and other variables* (Technical Report, Contract No. N61339-93-C-0031 with Naval Air Warfare Center). Orlando, FL: Naval Training Systems Center.
318	Kennedy, R. S. (1993). *Investigation and analysis of simulator sickness* (Final Report, Contract No. EERS-VN-C-92-025). Vienna, VA: EER Systems Corporation. Kennedy, R. S., Berbaum, K. S., & Lilienthal, M. G. (1997). Disorientation and postural ataxia following flight simulation. *Aviation, Space, and Environmental Medicine, 68*(1), 13-17.
Validation Study	Kennedy, R. S., Jones, S. A., & Smith, M. G. (1992). *The effects of platform motion on the incidence of simulator sickness in two rotary-wing flight trainers* (Final Report No. NTSC TR92-017). Orlando, FL: Naval Training Systems Center. Kennedy, R. S., Lane, N. E., & Braun, C. C. (1993). *A longitudinal field investigation of simulator sickness in the CH-46E/CH-53E flight trainers: Measurement issues related to usage of Image equipment and other variables* (Technical Report, Contract No. N61339-93-C-0031 with Naval Air Warfare Center). Orlando, FL: Naval Training Systems Center.

VR Studies

Study Number	Reference
101	Kolasinski, E. M. (1996). *Prediction of simulator sickness in a virtual environment*. Unpublished doctoral dissertation, University of Central Florida, Orlando. Kennedy, R. S., Jones, M. B., Stanney, K. M., Ritter, A. D., & Drexler, J. M. (1996). *Human factors safety testing for virtual environment mission-operation training* (Final Report, Contract No. NAS9-19482). Houston, TX: NASA Johnson Space Center.
102	Rich, C. J., & Braun, C. C. (1996). Assessing the impact of control and sensory compatibility on sickness in virtual environments. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 1122-1125). Santa Monica, CA: Human Factors & Ergonomics Society.

103	Bliss, J. P., Tidwell, P. D., Loftin, R. B., Johnston, B. E., Lyde, C. L., & Weathington, B. (1996). *An experimental evaluation of virtual reality for training teamed navigation skills* (Technical Report No. 96-01). Houston, TX: University of Houston, Virtual Environment Technology Laboratory.
104	Kennedy, R. S., Stanney, K. M., Dunlap, W. P., & Jones, M. B. (1996). *Virtual environment adaptation assessment test battery* (Final Report No. NASA1-96-1, Contract No. NAS9-19453). Houston, TX: NASA Johnson Space Center.
105	Bailey, J. H. (1994). *Spatial knowledge acquisition in a virtual environment*. Unpublished doctoral dissertation, University of Central Florida, Orlando. Bailey, J. H., & Witmer, B. G. (1994). Learning and transfer of spatial knowledge in a virtual environment. *Proceedings of the Human Factors & Ergonomics Society 38th Annual Meeting* (pp. 1158-1162). Santa Monica, CA: Human Factors and Ergonomics Society. Lampton, D. R., Kolasinski, E. M., Knerr, B. W., Bliss, J. P., Bailey, J. H., & Witmer, B. G. (1994). Side effects and aftereffects of immersion in virtual environments. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 1154-1157). Santa Monica, CA: Human Factors and Ergonomics Society. Lampton D. R., McDonald, D. P., Singer, M., & Bliss, J. P. (1995). Distance estimation in virtual environments. *Proceedings of the Human Factors and Ergonomics Society 39 Annual Meeting* (pp. 1268-1272). Santa Monica, CA: Human Factors and Ergonomics Society. Singer, M. J., Ehrlich, J., Cinq-Mars, S., & Papin, J. (1995, December). *Task performance in virtual environments: Stereoscopic versus monoscopic displays and head-coupling* (ARI Technical Report No. 1034). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
106	Bailey, J. H. (1994). *Spatial knowledge acquisition in a virtual environment*. Unpublished doctoral dissertation, University of Central Florida, Orlando. Bailey, J. H., & Witmer, B. G. (1994). Learning and transfer of spatial knowledge in a virtual environment. *Proceedings of the Human Factors & Ergonomics Society 38th Annual Meeting* (pp. 1158-1162). Santa Monica, CA: Human Factors and Ergonomics Society. Singer, M. J., Ehrlich, J., Cinq-Mars, S., & Papin, J. (1995, December). *Task performance in virtual environments: Stereoscopic versus monoscopic displays and head-coupling* (ARI Technical Report No. 1034). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
107 - 110	Champney, R. K. (2003). *Recovery from virtual environment exposure: Assessment methods, expected time-course of symptoms and potential readaptation mechanisms*. Unpublished master's thesis, University of Central Florida, Orlando.

	Stanney, K. M., Kingdon, K. S., Graeber, D., & Kennedy, R. S. (2002). Human performance in immersive virtual environments: Effects of exposure duration, user control, and scene complexity. *Human Performance, 15*(4), 339-366.
111 - 112	Meehan, M., Razzaque, S., Whitton, M., & Brooks, F. (2003). Effects of latency on presence in stressful virtual environments. *Proceedings of the IEEE Virtual Reality 2003* (pp. 141-148). Los Alamitos, CA: IEEE Computer Society Press.
113	Fidopiastis, C. M. (2006). *Virtual environment assessment and design for cognitive rehabilitation applications*. Unpublished doctoral dissertation, University of Central Florida, Orlando.
650	Lampton D. R., McDonald, D. P., Singer, M., & Bliss, J. P. (1995). Distance estimation in virtual environments. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1268-1272). Santa Monica, CA: Human Factors and Ergonomics Society. Witmer, B. G., & Sadowski, W. J., Jr. (1998). Nonvisually guided locomotion to a previously viewed target in real and virtual environments. *Human Factors*, *40*(3), 478-488.
651	Kline, P. B., & Witmer, B. G. (1996). Distance perception in virtual environments: Effects of field of view and surface texture at near distances. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 1112-1116). Santa Monica, CA: Human Factors and Ergonomics Society. Witmer, B. G., Bailey, J. H., Knerr, B. W., & Parsons, K. C. (1996). Virtual spaces and real world places: Transfer of route knowledge. *International Journal of Human-Computer Studies, 45*, 413-428. Witmer, B. G., & Kline, P. B. (1988). Judging perceived and traversed distance in virtual environments. *Presence*, *7*(2), 144-167. Witmer, B. G., & Sadowski, W. J., Jr. (1998). Nonvisually guided locomotion to a previously viewed target in real and virtual environments. *Human Factors*, *40*(3), 478-488.
725	Razzaque, S. (2005). *Redirected walking*. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill, NC. Razzaque, S., Swapp, D., Slater, M., Whitton, M. C., & Steed, A. (2002). Redirected walking in place. In S. Müller & W. Stüzlinger (Eds.), *Proceedings of Eighth Eurographics Workshop on Virtual Environments* (pp. 123-130). New York: ACM Press.
Validation Study	Zimmons, P. (2004). *The influence of lighting quality on presence and task performance in virtual environments*. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill, NC.

APPENDIX E:
REFERENCES FOR OTHER SIMULATOR AND VIRTUAL REALITY STUDIES AVAILABLE FOR INCLUSION IN THE SICKNESS DATABASE

Bidwell, L., & Knutson, K. (2005). Research conducted at the National Institute of Neurological Disorders and Stroke (NINDS), National Institute of Health.

Cohn, J., Muth, E. R., Schmorrow, D., Brendley, K., & Hillson, R. (2003). Reducing negative effects from virtual environments: Implications for just-in-time training. *RTO Meeting Proceedings 86: Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures* (RTO-MP-086; pp. 38-1 – 38-9). Neuilly-Sur-Seine, France: Research and Technology Organisation/North Atlantic Treaty Organisation (RTO/NATO).

Draper, M. H., Ruff, H. A., & LaFleur, T. (2001). The effects of camera control and display configuration on teleoperated target search tasks. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1872-1876). Santa Monica, CA: Human Factors and Ergonomics Society.

Gower, D. W., Jr., & Fowlkes, J. (1989, September). *Simulator sickness in the UH-60 (Black Hawk) flight simulator* (Final Report, USAARL Report No. 89-25). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Gower, D. W., Jr., Lilienthal, M. G., Kennedy, R. S., Fowlkes, J. E., & Baltzley, D. R. (1987, November). *Simulator sickness in the AH-64 Apache Combat simulator* (Final Report, USAARL Report No. 88-1). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Kennedy, R. S., Berbaum, K. S., Dunlap, W. P., & Smith, M. G. (1995). *Correlating visual scene elements with simulator sickness incidence: Hardware and software development* (Phase II Final Report, Contract No. N00019-92-C-0157). Washington, DC: Naval Air Systems Command.

Kuntz, L. A. (1996). *Body sway in response to retinal disparities simulating ocular torsion*. Unpublished doctoral dissertation, University of Florida, Gainesville.

Lange, B. S. (2006). *Pain and anxiety associated with minor medical procedures in paediatric clients and their parents: Is virtual reality a useful distraction technique?* Unpublished doctoral dissertation, University of South Australia, Adelaide.

Lange, B. S., Williams, M. T., Fulton, I., & Jones, S. (2005). *Subjective, physiological and adverse effects of the virtual reality ARQuake games compared to watching an animated movie in children aged 6-17 years.* Paper presented at Cybertherapy 2005: A Decade of VR, Basel, Switzerland, June 4-10.

Lilienthal, M. G., Fowlkes, J. E., Kennedy, R. S., Tabler, R. E., & Dutton, B. (1987, September). *Preliminary report on the in-plant evaluation of simulator sickness and related human factors engineering issues in the TH-57C training device*. Oak Ridge, TN: Martin Marietta Energy Systems, Inc.

Muth, E. R., & Lawson, B. (2003). Using flight simulators aboard ships: Human side effects of an optimal scenario with smooth seas. *Aviation, Space, and Environmental Medicine, 74*(5), 497-505.

Smyth, C. C. (2002, May). *Detecting targets from a moving vehicle with a head-mounted display and sound localization* (Technical Report No. ARL-TR-2703). Aberdeen Proving Ground, MD: Army Research Laboratory.

Smyth, C. C., Gombash, J. M., & Burcham, P. M. (2001, June). *Indirect vision driving with fixed flat panel displays for near-unity, wide, and extended fields of camera view* (Technical Report No. ARL-TR-2511). Aberdeen Proving Ground, MD: Army Research Laboratory.

APPENDIX F:
SIGNIFICANT COMPARISONS ON TOTAL SEVERITY SCORE FOR SIMULATOR AND VR STUDIES

Simulator Studies

Study Pair	Simulator Type	Mean Difference	Significance
202, 204	Driving, Driving	28.405	$p < .001$
202, 203	Driving, Driving	26.123	$p < .001$
201, 202	Driving, Driving	18.404	$p = .034$
201, 204	Driving, Driving	10.001	$p = .036$
202, 315	Driving, Rotary	32.037	$p < .001$
202, 309	Driving, Rotary	31.017	$p < .001$
202, 317	Driving, Rotary	29.900	$p < .001$
202, 305	Driving, Rotary	28.906	$p < .001$
202, 314	Driving, Rotary	28.048	$p < .001$
202, 311	Driving, Rotary	27.139	$p < .001$
202, 313	Driving, Rotary	26.230	$p < .001$
202, 310	Driving, Rotary	22.659	$p = .002$
202, 312	Driving, Rotary	20.967	$p = .003$
201, 315	Driving, Rotary	13.633	$p = .006$
201, 309	Driving, Rotary	12.613	$p = .001$
201, 305	Driving, Rotary	10.502	$p = .022$
204, 312	Driving, Rotary	7.438	$p = .083$
202, 318	Driving, Fixed	37.179	$p < .001$
202, 304	Driving, Fixed	36.712	$p < .001$
202, 306	Driving, Fixed	34.281	$p < .001$
202, 316	Driving, Fixed	33.809	$p < .001$
202, 308	Driving, Fixed	31.289	$p < .001$
202, 302	Driving, Fixed	31.236	$p < .001$
202, 307	Driving, Fixed	30.982	$p < .001$
202, 303	Driving, Fixed	19.363	$p = .134$
201, 318	Driving, Fixed	18.775	$p < .001$
201, 304	Driving, Fixed	18.308	$p < .001$
201, 306	Driving, Fixed	15.877	$p < .001$
201, 316	Driving, Fixed	15.404	$p < .001$
201, 308	Driving, Fixed	12.885	$p = .006$
201, 302	Driving, Fixed	12.831	$p = .009$
201, 307	Driving, Fixed	12.578	$p = .004$
203, 318	Driving, Fixed	11.057	$p = .070$
204, 318	Driving, Fixed	8.775	$p = .016$
204, 304	Driving, Fixed	8.307	$p = .134$
312, 318	Rotary, Fixed	16.213	$p < .001$
304, 312	Fixed, Rotary	15.745	$p < .001$

310, 318	Rotary, Fixed	14.520	$p < .001$
304, 310	Fixed, Rotary	14.053	$p = .005$
306, 312	Fixed, Rotary	13.314	$p < .001$
312, 316	Rotary, Fixed	12.842	$p = .001$
306, 310	Fixed, Rotary	11.622	$p = .019$
310, 316	Rotary, Fixed	11.149	$p = .045$
313, 318	Rotary, Fixed	10.949	$p = .068$
308, 312	Fixed, Rotary	10.322	$p = .016$
302, 312	Fixed, Rotary	10.269	$p = .026$
311, 318	Rotary, Fixed	10.040	$p = .013$
307, 312	Fixed, Rotary	10.016	$p = .009$
304, 311	Fixed, Rotary	9.573	$p = .087$
314, 318	Rotary, Fixed	9.132	$p = .098$
305, 318	Rotary, Fixed	8.274	$p = .033$
303, 318	Fixed, Fixed	17.817	$p = .031$
303, 304	Fixed, Fixed	17.349	$p = .050$
303, 306	Fixed, Fixed	14.918	$p = .132$
312, 315	Rotary, Rotary	11.070	$p = .018$
309, 312	Rotary, Rotary	10.050	$p = .001$
305, 312	Rotary, Rotary	7.939	$p = .048$

VR Studies

Study Pair	VR Type	Mean Difference	Significance
102, 113	HMD, HMD	42.698	$p = .021$
102, 112	HMD, HMD	42.099	$p = .023$
102, 111	HMD, HMD	40.906	$p = .029$
105, 113	HMD, HMD	26.399	$p < .001$
105, 112	HMD, HMD	25.800	$p < .001$
105, 111	HMD, HMD	24.607	$p < .001$
110, 113	HMD, HMD	24.601	$p < .001$
110, 112	HMD, HMD	24.002	$p < .001$
109, 113	HMD, HMD	23.285	$p < .001$
110, 111	HMD, HMD	22.809	$p < .001$
109, 112	HMD, HMD	22.686	$p < .001$
109, 111	HMD, HMD	21.493	$p < .001$
108, 113	HMD, HMD	20.050	$p < .001$
108, 112	HMD, HMD	19.450	$p < .001$
106, 112	HMD, HMD	19.285	$p = .049$

108, 111	HMD, HMD	18.257	$p < .001$
101, 105	HMD, HMD	16.804	$p = .022$
107, 113	HMD, HMD	15.639	$p < .001$
107, 112	HMD, HMD	15.040	$p < .001$
101, 110	HMD, HMD	15.006	$p < .001$
107, 111	HMD, HMD	13.847	$p = .001$
101, 109	HMD, HMD	13.690	$p = .001$
101, 108	HMD, HMD	10.454	$p = .030$
101, 112	HMD, HMD	8.996	$p = .043$
105, 651	HMD, BOOM	18.880	$p = .005$
110, 651	HMD, BOOM	17.082	$p < .001$
109, 651	HMD, BOOM	15.766	$p < .001$
108, 651	HMD, BOOM	12.531	$p = .005$
102, 725	HMD, CAVE	45.735	$p = .012$
105, 725	HMD, CAVE	29.436	$p < .001$
110, 725	HMD, CAVE	27.638	$p < .001$
109, 725	HMD, CAVE	26.322	$p < .001$
108, 725	HMD, CAVE	23.086	$p < .001$
106, 725	HMD, CAVE	22.921	$p = .008$
107, 725	HMD, CAVE	18.676	$p < .001$
101, 725	HMD, CAVE	12.632	$p < .001$
651, 725	BOOM, CAVE	10.555	$p = .007$

LIST OF REFERENCES

Allen, R. C., Singer, M. J., McDonald, D. P., & Cotton, J. E. (2000). Age differences in a virtual reality entertainment environment: A field study. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1-542 – 1-545). Santa Monica, CA: Human Factors and Ergonomics Society.

Allison, R. S., Harris, L. R., Jenkin, M., Jasiobedzka, U., & Zacher, J. E. (2001). Tolerance of temporal delay in virtual environments. *Proceedings of the IEEE Virtual Reality 2001 International Conference* (pp. 247-254). New York: Institute of Electrical and Electronics Engineers, Inc.

Badiqué, E., Cavazza, M., Klinker, G., Mair, G., Sweeney, T., Thalmann, D., & Thalmann, N. M. (2002). Entertainment applications of virtual environments. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 1143-1166). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Baltzley, D. R., Kennedy, R. S., Berbaum, K. S., Lilienthal, M. G., & Gower, D. W. (1989). The time course of postflight simulator sickness symptoms. *Aviation, Space, and Environmental Medicine*, *60*(11), 1043-1048.

Bell, J. T., & Fogler, H. S. (1998). Virtual reality in chemical engineering education. *Proceedings of the American Society for Engineering Education (ASEE) North Central Section Meeting*. Available: http://www.vrupl.evl.uic.edu/vrichel/Papers/aseepap7.pdf

Bernasch, J., & Haenel, S. (1995). The BMW driving simulator used for the development of a driver-biased adaptive cruise control. *Proceedings of the Driving Simulation Conference "DSC'95"* (pp. 157-174). Montigny-le-Bretonneux, France: Neuf Associés.

Biocca, F. (1992). Will simulator sickness slow down the diffusion of virtual environment technology? *Presence, 1*(3), 334-343.

Blade, R. A., & Padgett, M. (2002). Virtual environments standards and terminology. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 15-27). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Boff, K. R., & Lincoln, J. E. (Eds.). (1988a). Factors affecting sensitivity to flicker. *Engineering data compendium: Human perception and performance, Vols. 1* (pp. 166-167). Wright-Patterson Air Force Base, OH: Harry G. Armstrong, Aerospace Medical Research Laboratory.

Boff, K. R., & Lincoln, J. E. (Eds.). (1988b). Flicker sensitivity: Effect of flicker frequency and luminance level. *Engineering data compendium: Human perception and performance, Vol. 1* (pp. 170-171). Wright-Patterson Air Force Base, OH: Harry G. Armstrong, Aerospace Medical Research Laboratory.

Boff, K. R., & Lincoln, J. E. (Eds.). (1988c). Flicker sensitivity: Effect of target size. *Engineering data compendium: Human perception and performance, Vol. 1* (pp. 178-179). Wright-Patterson Air Force Base, OH: Harry G. Armstrong, Aerospace Medical Research Laboratory.

Boff, K. R., & Lincoln, J. E. (Eds.). (1988d). Flicker thresholds for various cathode ray tube phosphors. *Engineering data compendium: Human perception and performance, Vol. 3* (pp. 2258-2259). Wright-Patterson Air Force Base, OH: Harry G. Armstrong, Aerospace Medical Research Laboratory.

Boulanger, O., & Chevennement, J. (1995). Analytical and application experiments: Two necessary approaches for the driving simulators validity. *Proceedings of the Driving Simulation Conference "DSC'95"* (pp.25-39). Montigny-le-Bretonneux, France: Neuf Associés.

Bowman, D. A., Datey, A., Ryu, Y. S., Farooq, U., & Vasnaik, O. (2002). Empirical comparison of human behavior and performance with different display devices for virtual environments. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 2134-2138). Santa Monica, CA: Human Factors and Ergonomics Society.

Boynton, R. M., & Onley, J. W. (1962). A critique of the special status assigned by Brindley to "psychophysical linking hypotheses" of "Class A". *Vision Research, 2*, 383-390.

Brindley, G. S. (1960). *Physiology of the retina and the visual pathway*. Baltimore: Edward Arnold, London, and William and Wilkins.

Bridgeman, B. (1995). Direction constancy in rapidly refreshed video displays. *Journal of Vestibular Research, 5*(6), 393-398.

Brown, D. J., Kerr, S. J., & Bayon, V. (1998). The development of the Virtual City: A user centered approach. *Proceedings of the 2nd European Conference on Disability, Virtual Reality, and Associated Technology* (pp. 11-15). Skovde, Sweden: University of Reading.

Bryson, S. (2002). Information visualization in virtual environments. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 1101-1118). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Canaras, S. A., Gentner, F. C., Schopper, A. W. (1995, July). *Virtual reality (VR) training* (Final Report No. CSERIAC-RA-95-009). Wright Patterson Air Force Base, OH: Crew System Ergonomics Information Analysis Center.

Casali, J. G., & Wierwille, W. W. (1980). The effects of various design alternatives on moving-base driving simulator discomfort. *Human Factors, 22*(6), 741-756.

Cheung, B. S. K., Howard, I. P., & Money, K. E. (1991). Visually-induced sickness in normal and bilaterally labyrinthine-defective subjects. *Aviation, Space, and Environmental Medicine, 62*(6), 527-531.

Cobb, S., Neale, H., Crosier, J., & Wilson, J. R. (2002). Development and evaluation of virtual environments for education. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 911-936). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cobb, S. V. G., Neale, H. R., & Reynolds, H. (1998). Evaluation of virtual learning environments. *Proceedings of the 2nd European Conference on Disability, Virtual Reality, and Associated Technology* (pp. 17-23). Skovde, Sweden: University of Reading.

Cobb, S. V. G., Nichols, S., Ramsey, A., & Wilson, J. R. (1999). Virtual reality-induced symptoms and effects (VRISE). *Presence, 8*(2), 169-186.

Cosby, P. C. (1993). *Methods in behavioral research* (5th ed.). Mountain View, CA: Mayfield Publishing Company.

Covault, C. (1998). Virtual reality utilized in Station, Shuttle Ops. *Aviation Week & Space Technology, 149*(13), 74-76.

Crampton, G. H. (1990). Neurophysiology of motion sickness. In G. H. Crampton (Ed.), *Motion and space sickness* (pp. 29-42). Boca Raton, FL: CRC Press, Inc.

Crowley, J. S. (1987). Simulator sickness: A problem for Army aviation. *Aviation, Space, and Environmental Medicine, 58*(4), 355-357.

Curry, R., Artz, B., Cathey, L., Grant, P., & Greenberg, J. (2002). Kennedy SSQ results: Fixed- vs. motion-base Ford simulators. *Proceedings of the Driving Simulation Conference "DSC2002"* (pp. 289-299).

Department of the Navy. (2004, March 1). *NATOPS General Flight and Operating Instructions (OPNAVINST 3710.7T), Section 8.3.2.17: Simulator sickness* (p. 8-10). Washington, DC: Author.

DeWit, G. (1957). Acquired sensitivity to seasickness after an influenza infection. *Practica Otorhinolaryngologica, 19*, 579-586.

DiZio, P., & Lackner, J R. (1992). Spatial orientation, adaptation, and motion sickness in real and virtual environments. *Presence, 1*(3), 319-328.

DiZio, P., & Lackner, J R. (1997). Circumventing side effects of immersive virtual environments. In M.J. Smith, G. Salvendy, & R. J. Koubek (Eds.), *Design of computing systems: Social and ergonomic considerations* (pp. 893-896). Amsterdam: Elsevier.

Dobie, T., McBride, D., Dobie, T., Jr., & May, J. (2001). The effects of age and sex on susceptibility to motion sickness. *Aviation, Space, and Environmental Medicine, 72*(1), 13-20.

Draper, M. H., Viirre, E. S., Furness, T. A., & Gawron, V. J. (2001). Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors, 43*(1), 129-146.

Drexler, J. M., Kennedy, R. S., & Compton, D. E. (2004). *Comparison of sickness profiles from simulator and virtual environment devices: Implications of engineering features*. Paper presented at the 2004 Driving Simulation Conference Europe "DSC 2004", September 8-10, Paris, France.

Durlach, N. I., & Mavor, A. S. (Eds.). (1995). *Virtual reality: Scientific and technological challenges*. Washington, DC: National Academy Press.

Ebenholtz, S. M. (1988). *Sources of asthenopia in Navy flight simulators* (Final Report, Accession No. AD-A212699). Alexandria, VA: Defense Logistics Agency, Defense Technical Information Center.

Ebenholtz, S. M. (1992). Motion sickness and oculomotor systems in virtual environments. *Presence, 1*(3), 302-305.

Ebenholtz, S. M. (2001). *Oculomotor systems and perception*. Cambridge, UK: Cambridge University Press.

Flanagan, M. B., May, J. G., & Dobie, T. G. (2005). Sex differences in tolerance to visually-induced motion sickness. *Aviation, Space, and Environmental Medicine, 76*(7), 642-646.

Förstberg, J., & Ledin, T. (1996). *Discomfort caused by low-frequency motions: A literature survey of hypotheses and possible causes of motion sickness* (TRITA-FKT Report 1996:39). Stockholm: Swedish National Road and Transportation Research Institute.

Fowlkes, J., Durlach, P. J., Drexler, J. M., Daly, J., Alberdeston, R., & Metevier, C. (2002). Optimizing haptics perceptions for advanced Army training systems: Impacts on performance. *Proceedings of the 23rd Annual Army Science Conference* [On-line]. Available: http://www.asc2002.com/23rdASC/manuscripts/M/MO-01.pdf

Frank, L. H., Casali, J. H., & Wierwille, W. W. (1988). Effects on visual display and motion system delays on operator performance and uneasiness in a driving simulator. *Human Factors, 30*(2), 201-217.

Garris-Reif, R., & Franz, T. M. (1995). Simulator sickness and human task performance in conventional and virtual environments. In A.C. Bittner & P.C. Champney (Eds.), *Advances in Industrial Ergonomics and Safety VII* (pp. 219-223). London: Taylor & Francis.

Gianaros, P. J., Muth, E. R., Mordkoff, J. T., Levine, M. E., & Stern, R. M. (2001). A questionnaire for the assessment of the multiple dimensions of motion sickness. *Aviation, Space, and Environmental Medicine, 72*(2), 115-119.

Gottesman, I.I., & Gould, T.D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry, 160*, 636-645.

Gower, D. W., Lilienthal, M. G., Kennedy, R. S., & Fowlkes, J. E. (1987, September). Simulator sickness in U.S. Army and Navy fixed- and rotary-wing flight simulators. In *Conference Proceedings of the AGARD Medical Panel Symposium on Motion Cues in Flight Simulation and Simulator Induced Sickness* (AGARD-CP-433; pp. 8.1 - 8.20). Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development.

Graeber, D. A., & Stanney, K. M. (2002). Gender differences in visually induced motion sickness. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 2109-2113). Santa Monica, CA: Human Factors and Ergonomics Society.

Griffin, M J. (1991). Sea sickness. *In AGARD Lecture Series 175: Motion sickness: Significance in aerospace operations and prophylaxis* (AGARD-LS-175; pp. 7.1 – 7.20). Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development.

Guignard, J. C., & McCauley, M. E. (1990). The accelerative stimulus for motion sickness. In G. H. Crampton (Ed.), *Motion and space sickness* (pp. 123-152). Boca Raton, FL: CRC Press, Inc.

Häkkinen, J., Vuori, T., & Puhakka, M. (2002). *Postural stability and sickness symptoms after HMD use*. Paper presented at the IEEE International Conference on Systems, Man and Cybernetics. Available online: http://www.nokia.com/nokia

Harm, D. L. (1990). Physiology of motion sickness symptoms. In G. H. Crampton (Ed.), *Motion and space sickness* (pp. 153-177). Boca Raton, FL: CRC Press, Inc.

Harm, D. L. (2002). Motion sickness neurophysiology, physiological correlates, and treatment. In K.M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 637-661). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hettinger, L. J. (2002). Illusory self-motion in virtual environments. In K.M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 471-491). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hettinger, L. J., Berbaum, K. S., Kennedy, R. S., Dunlap, W. P., & Nolan, M. D. (1990). Vection and simulator sickness. *Military Psychology, 2*(3), 171-181.

Hettinger, L. J., Lilienthal, M. G., Kennedy, R. S., Berbaum, K. S., & Hooper, J. M. (1987, April). *Addressing the problem of simulator sickness: A forum for expert recommendations*. Paper presented at the Aviation Psychology Conference, Columbus, OH.

Hettinger, L. J., Nolan, M. D., Kennedy, R. S., Berbaum, K. S., & Schnitzius, K. P., & Edinger, K. M. (1987). Visual display factors contributing to simulator sickness. *Proceedings of the 31st Annual Meeting of the Human Factors Society* (pp. 497-501). Santa Monica, CA: Human Factors Society.

Hettinger, L. J., & Riccio, G. E. (1992). Visually induced motion sickness in virtual environments. *Presence, 1*(3), 306-310.

Hix, D., Swan, J. E., Gabbard, M., McGee, M., Durbin, J., & King, T. (1999). User-centered design and evaluation of a real-time battlefield visualization virtual environment. In L. Rosenblum, P. Astheimer, & D. Teichmann (Eds.), *Proceedings IEEE Virtual Reality '99* (pp. 96-103). Los Alamitos, CA: IEEE Computer Society Press.

Hodges, L. F., Rothbaum, B. O., Kooper, R., Opdyke, D., Meyer, T. C., North, M. M, de Graaff, J. J., & Williford, J. S. (1995). Virtual environments for treating the fear of heights. *IEEE Computer, 28*(7), 27-34.

Howarth, P. A., & Costello, P. J. (1996, May). *Visual effects of immersion in virtual environments: Interim results from the U.K. Health and Safety Executive Study*. Paper presented at Society for Information Display International Symposium, San Diego, CA.

Jaeger, B. K., & Mourant, R. R. (2001). Comparison of simulator sickness using static and dynamic walking simulators. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1896-1900). Santa Monica, CA: Human Factors and Ergonomics Society.

Johnson, W. H., Sunahara, F. A., & Landolt, J. P. (1993). Motion sickness, vascular changes accompanying pseudo-Coriolis-induced nausea. *Aviation, Space, and Environmental Medicine, 64*(5), 367-370.

Jones, S. A. (1998). *Effects of restraint on vection and simulator sickness*. Unpublished doctoral dissertation, University of Central Florida, Orlando.

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.

Kennedy, R. S. (1975). Motion sickness questionnaire and field independence scores as predictors of success in naval aviation training. *Aviation, Space, and Environmental Medicine, 46*, 1349-1352.

Kennedy, R. S. (1996). *Analysis of simulator sickness data* (Technical Report, Contract No. N61339-91-D-0004 with Enzian Technology, Inc). Orlando, FL: Naval Air Warfare Center, Training Systems Division.

Kennedy, R. S., Allgood, G. O., Van Hoy, B. W., & Lilienthal, M. G. (1987, June). Motion sickness symptoms and postural changes following flights in motion-based flight trainers. *Journal of Low Frequency Noise and Vibration, 6*(4), 147-154.

Kennedy, R. S., Berbaum, K. S., Allgood, G. O., Lane, N. E., & Lilienthal, M. G., & Baltzley, D. R. (1988). Etiological significance of equipment features and pilot history in simulator sickness. *AGARD Conference Proceedings No. 433: Motion Cues in Flight Simulation and Simulator Induced Sickness* (pp. 1.1-1.22). Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development.

Kennedy, R. S., Berbaum, K. S., Dunlap, W. P., & Hettinger, L. J. (1996). Developing automated methods to quantify the visual stimulus for cybersickness. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 1126-1130). Santa Monica, CA: Human Factors and Ergonomics Society.

Kennedy, R. S., Berbaum, K. S., Dunlap, W. P., & Smith, M. G. (1995, October). *Correlating visual scene elements with simulator sickness incidence: Hardware and software development* (Phase II Final Report, Contract No. N00019-92-C-0157). Washington, DC: Naval Air Systems Command.

Kennedy, R. S., Berbaum, K. S., & Lilienthal, M. G. (1997). Disorientation and postural ataxia following flight simulation. *Aviation, Space, and Environmental Medicine, 68*(1), 13-17.

Kennedy, R. S., Berbaum, K. S., Lilienthal, M. G., Dunlap, W. P., Mulligan, B. E., & Funaro, J. F. (1987). *Guidelines for alleviation of simulator sickness symptomatology* (Final Report No. NAVTRASYSCEN TR-87-007). Orlando, FL: Naval Training Systems Center.

Kennedy, R. S., Berbaum, K. S., & Smith, M. G. (1993). Methods for correlating visual scene elements with simulator sickness incidence. *Proceedings of the 37th Annual Meeting of the Human Factors Society* (pp. 1252-1256). Santa Monica, CA: Human Factors and Ergonomics Society.

Kennedy, R. S., Drexler, J. M., & Compton, D. E. (1997). Simulator sickness and other aftereffects: Implications for the design of driving simulators. *Proceedings of the Driving Simulation Conference (DSC'97*; pp. 115-123). Paris, France: ETNA.

Kennedy, R. S., Drexler, J. M., Compton, D. E., Stanney, K. M., Lanham, D. S., & Harm, D. L. (2003). Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: Similarities and differences. In L.J. Hettinger & M.W. Haas (Eds.), *Virtual and adaptive environments: Applications, implications, and human performance* (pp. 247-278). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kennedy, R. S., Drexler, J. M., Stanney, K. M., & Harm, D. L. (1997). Configural scoring of the self-report of symptoms in different motion sickness environments: Normative data and comparison with other scoring systems. Published in *Abstracts of the International Workshop on Motion Sickness: Medical and Human Factors* (pp.78-82).

Kennedy, R. S., Dunlap, W. P., & Fowlkes, J. E. (1990). Prediction of motion sickness susceptibility. In G. H. Crampton (Ed.), *Motion and space sickness* (pp. 179-215). Boca Raton, FL: CRC Press, Inc.

Kennedy, R. S., Dunlap, W. P., Jones, M. B., & Stanney, K. M. (1996). *Screening users of virtual reality systems for after-effects such as motion sickness and balance problems* (Final Report No. NSF1-96-4, Grant No. DMI-9561266). Arlington, VA: National Science Foundation.

Kennedy, R. S., & Fowlkes, J. E. (1992). Simulator sickness is polygenic and polysymptomatic: Implications for research. *International Journal of Aviation Psychology*, *2*(1), 23-38.

Kennedy, R. S., Fowlkes, J. E., Berbaum, K. S., & Lilienthal, M. G. (1992). Use of a motion sickness history questionnaire for prediction of simulator sickness. *Aviation, Space, and Environmental Medicine, 63*, 588-93.

Kennedy, R. S., Fowlkes, J. E., & Hettinger, L. J. (1989). *Review of simulator sickness literature* (Technical Report No. NTSC TR89-024). Orlando, FL: Naval Training Systems Center.

Kennedy, R. S., Fowlkes, J. E., & Lilienthal, M. G. (1993). Postural and performance changes following exposures to flight simulators. *Aviation, Space, and Environmental Medicine*, *64*, 912-920.

Kennedy, R. S., & Frank, L H. (1985). *A review of motion sickness with special reference to simulator sickness* (Technical Report No. NAVTRAEQUIPCEN 81-C-0105-16). Orlando, FL: Naval Training Equipment Center.

Kennedy, R. S., Frank, L. H., & McCauley, M. E. (1985). *Simulator sickness: Reaction to a transformed perceptual world. II. Sourcebook and suggested readings* (Technical Report No. NAVTRAEQUIPCEN 81-C-0105-7). Orlando, FL: Naval Training Equipment Center. (AD No. A210 512)

Kennedy, R. S., & Graybiel, A. (1965). *The Dial test: A standardized procedure for the experimental production of canal sickness symptomatology in a rotating environment* (Report No. 113, NSAM 930). Pensacola, FL: Naval School of Aerospace Medicine.

Kennedy R. S., Graybiel, A., McDonough, R. C., & Beckwith, F. D. (1968). Symptomatology under storm conditions in the North Atlantic in control subjects and in persons with bilateral labyrinthine defects. *Acta Otolaryngologica, 66*, 533-540.

Kennedy, R. S., Hettinger, L. J., & Lilienthal, M. G. (1990). Simulator sickness. In G. H. Crampton (Ed.), *Motion and space sickness* (pp. 317-341). Boca Raton, FL: CRC Press, Inc.

Kennedy, R. S., Jones, M. B., & Dunlap, W. P. (1996). A predictive model of simulator sickness: Applications for virtual reality [Abstract]. *Aviation, Space, and Environmental Medicine, 67*(7), 672.

Kennedy, R. S., Jones, M. B., Lilienthal, M. G., & Harm, D. L. (1994). Profile analysis of after-effects experienced during exposure to several virtual reality environments. In *Conference Proceedings of the AGARD Medical Panel Symposium on Virtual Interface: Research & Applications* (AGARD-CP-541; pp. 2.1-2.9). Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development.

Kennedy, R. S., Jones, M. B., Stanney, K. M., Ritter, A., & Drexler, J. M. (1996). *Human factors safety testing for virtual environment mission-operations training* (Final Report No. NASA1-96-2, Contract No. NAS9-19482). Houston, TX: NASA Johnson Space Center.

Kennedy, R. S., Kennedy, K. E., & Bartlett, K. M. (2002). Virtual environments and product liability. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 543-553). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire (SSQ): A new method for quantifying simulator sickness. *International Journal of Aviation Psychology*, *3*(3), 203-220.

Kennedy, R. S., Lane, N. E., & Kuntz, L. A. (1987). Surrogate measures: A proposed alternative in human factors assessment of operational measures of performance. *Proceedings of the 1st Annual Workshop on Space Operations, Automation, & Robotics* (pp. 551-558). Houston, TX: NASA Lyndon B. Johnson Space Center.

Kennedy, R. S., Lane, N. E., Lilienthal, M. G., Berbaum, K. S., & Hettinger, L. J. (1992). Profile analysis of simulator sickness symptoms: Application to virtual environment systems. *Presence, 1*(3), 295-301.

Kennedy, R. S., Lanham, D. S., Drexler, J. M., Massey, C. J., & Lilienthal, M. G. (1997). A comparison of cybersickness incidences, symptom profiles, measurement techniques, and suggestions for further research. *Presence, 6*(6), 638-644.

Kennedy, R. S., Lanham, D. S., Massey, C. J., Drexler, J. M., & Lilienthal, M. G. (1995). Gender differences in simulator sickness incidence: Implications for military virtual reality systems. *SAFE Journal, 25*(1), 69-76.

Kennedy, R. S., & Lilienthal, M. G. (1994). Measurement and control of motion sickness aftereffects from immersion in virtual reality. *Proceedings of Virtual Reality and Medicine, The Cutting Edge* (pp. 111-119). New York: SIG-Advanced Applications, Inc.

Kennedy, R. S., Lilienthal, M. G., Berbaum, K. S., Baltzley, D. R., & McCauley, M. E. (1989). Simulator sickness in U.S. Navy flight simulators. *Aviation, Space, and Environmental Medicine*, *60*, 10-16.

Kennedy, R. S., & Smith, M. G. (1996, November). *A smart system to control stimulation for visually induced motion sickness* (Phase II Final Report No. NAS9-19106). Houston, TX: NASA Lyndon B. Johnson Space Center.

Kennedy, R. S., & Stanney, K. M. (1996). Postural instability induced by virtual reality exposure: Development of a certification protocol. *International Journal of Human-Computer Interaction, 8*(1), 25-47.

Kennedy, R. S., & Stanney, K. M. (1996). Virtual reality systems and products liability. *The Journal of Medicine and Virtual Reality, 1*(2), 60-64.

Kennedy, R. S., & Stanney, K. M. (1997). Aftereffects of virtual environment exposure: Psychometric issues. In M.J. Smith, G. Salvendy, & R.J. Koubek (Eds.), *Design of computing systems: Social and ergonomic considerations* (pp. 897-900). Amsterdam: Elsevier.

Kennedy, R. S., Stanney, K. M., & Dunlap, W. P. (2000). Duration and exposure to virtual environments: Sickness curves during and across sessions. *Presence, 9*(5), 463-472.

Kennedy, R. S., Stanney, K. M., Dunlap, W. P., & Jones, M. B. (1996). *Virtual environment adaptation assessment test battery* (Final Report No. NASA1-96-1, Contract No. NAS9-19453). Houston, TX: NASA Johnson Space Center.

Kennedy, R. S., Stanney, K. M., & Fernandez, E. (1999). *Six months residual after effects from a virtual reality entertainment system*. Unpublished manuscript.

Kennedy, R. S., Tolhurst, G. C., & Graybiel, A. (1965). *The effects of visual deprivation on adaptation to a rotating environment* (Report No. 106, NSAM 918). Pensacola, FL: Naval School of Aviation Medicine.

Kennedy, R. S., Turnage, J. J., & Lane, N. E. (1995). Development of surrogate measurement methodologies for workplace performance. In A.C. Bittner & P.C. Champney (Eds.), *Advances in Industrial Ergonomics & Safety VII* (pp. 485-492). London: Taylor & Francis.

Kennedy, R. S., Turnage, J. J., & Lane, N. E. (1997). Development of surrogate methodologies for operational performance measurement: Empirical studies. *Human Performance, 10*(3), 251-282.

Kenyon, R. V., & Afenya, M. B. (1995). Training in virtual and real environments. *Annals of Biomedical Engineering, 23*, 445-455.

Kingdon, K. S., Stanney, K. M., & Kennedy, R. S. (2001). Extreme responses to virtual environment exposure. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1906-1911). Santa Monica, CA: Human Factors and Ergonomics Society.

Knerr, B. W., Breaux, R., Goldberg, S. L., & Thurman, R. A. (2002). National defense. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 857-872). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kolasinski, E. M. (1995, May). *Simulator sickness in virtual environments* (ARI Technical Report No. 1027). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

La Viola, J. J., Jr. (2000). A discussion of cybersickness in virtual environments. *SIGCHI Bulletin, 32*(1), 47-56.

Lampton, D. R., Kraemer, R. E., Kolasinski, E. M., & Knerr, B. W. (1995, October). *An investigation of simulator sickness in a tank driver trainer* (ARI Report No. 1684). Orlando, FL: U.S. Army Research Institute for the Behavioral and Social Sciences, Simulator Systems Research Unit.

Lampton, D. R., Rodriguez, M. E., & Cotton, J. E. (2000). Simulator sickness symptoms during team training in immersive virtual environments. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1-530 – 1-533). Santa Monica, CA: Human Factors and Ergonomics Society.

Lane, N. E., & Kennedy, R. S. (1988). *A new method for quantifying simulator sickness: Development and application of the simulator sickness questionnaire (SSQ)* (Report No. EOTR 88-7). Orlando, FL: Essex Corporation.

Lane, N. E., Kennedy, R. S., & Jones, M. B. (1986). Overcoming unreliability in operational measures: The use of surrogate measure systems. *Proceedings of the 30th Annual Meeting of the Human Factors Society* (pp. 1398-1402). Dayton, OH: Human Factors Society.

Lathan, C. E., Tracey, M. R., Sebrechts, M. M., Clawson, D. M., & Higgins, G. A. (2002). Using virtual environments as training simulators: Measuring transfer. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 403-414). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lawson, B. D., Graeber, D. A., Mead, A. M., & Muth, E. R. (2002). Signs and symptoms of human syndromes associated with synthetic experiences. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 589-618). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lentz, J. M., & Collins, W. E. (1977). Motion sickness susceptibility and related behavioral characteristics in men and women. *Aviation, Space, and Environmental Medicine, 48*(4), 316-322.

Lerman, Y., Sadovsky, G., Goldberg, E., Kedem, R., Peritz, E., & Pines, A. (1993). Correlates of military tank simulator sickness. *Aviation, Space, and Environmental Medicine, 64*(7), 619-622.

Lin, J. J-W., Duh, H. B. L., Parker, D. E., Abi-Rached, H., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings of the IEEE Virtual Reality Conference 2002* (pp. 164-171). New York: Institute of Electrical and Electronics Engineers, Inc.

Magee, L. E. (1995, March). *Virtual Reality Simulator (VRS) for training ship handling skills*. Paper presented at the NATO/OCTAN Research Study Group 16 "Advanced Technologies Applied to Training Design" Workshop: Virtual Environments Training's Future?, Portsmouth, England.

Malaterre, G. (1995). Comparisons between simulation and actual driving situations: Some experiments. *Proceedings of the Driving Simulation Conference "DSC95"* (pp. 59-76). Montigny-le-Bretonneux, France: Neuf Associés.

May, J. G., & Badcock, D. R. (2002). Vision and virtual environments. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 29-63). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McCauley, M. E. (Ed.). (1984). *Simulator sickness: Proceedings of a workshop*. Washington, DC: National Academy Press.

McCauley, M. E., & Sharkey, T. J. (1992). Cybersickness: Perception of self-motion in virtual environments. *Presence, 1*, 311-318.

McCauley-Bell, P. R. (2002). Ergonomics in virtual environments. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 807-826). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mendenhall, W., & Sincich, T. (1995). *Statistics for engineering and the sciences* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Miller, J. C., Sharkey, T. J., Graham, G. A., & McCauley, M. E. (1993). Autonomic physiological data associated with simulator discomfort. *Aviation, Space, and Environmental Medicine, 64*(9), 813-819.

Miller, J. W., & Goodson, J. E. (1960). Motion sickness in a helicopter simulator. *Aerospace Medicine*, *31*(3), 204-212.

Miller, K. E., & Muth, E. R. (2004). Efficacy of acupressure and acustimulation bands for the prevention of motion sickness. *Aviation, Space, and Environmental Medicine, 75*(3), 227-234.

Moldenhauer, R. (1995). Driver training with simulation of traffic and the environment. *Proceedings of the Driving Simulation Conference "DSC95"* (pp. 385-394). Montigny-le-Bretonneux, France: Neuf Associés.

Mon-Williams, M., & Wann, J. P. (1998). Binocular virtual reality displays: When problems do and don't occur. *Human Factors, 40*(1), 42-49.

Mon-Williams, M., Wann, J. P., & Rushton, S. (1993). Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display. *Ophthalmic and Physiological Optics, 13*, 387-391.

Mon-Williams, M., Wann, J. P., & Rushton, S. (1995). Design factors in stereoscopic virtual-reality displays. *Journal of the SID, 3/4*, 207-210.

Money, K. E. (1970). Motion sickness. *Physiological Reviews, 50*(1), 1-31.

Money, K. E. (1990). Motion sickness and evolution. In G. H. Crampton (Ed.), *Motion and space sickness* (pp. 1-7). Boca Raton, FL: CRC Press, Inc.

Money, K. E., & Cheung, B. S. (1982). A mechanism for facilitation of the emetic response to poisons: The basis of motion sickness. *Proceedings of the Aerospace Medical Association Meeting* (pp. 140-141).

Money, K. E., & Cheung, B. S. (1983). Another function of the inner ear: Facilitation of the emetic response to poisons. *Aviation, Space, and Environmental Medicine*, *54*(3), 208-211.

Money, K. E., Lackner, J. R., & Cheung, R. S. K. (1996). The autonomic nervous system and motion sickness. In B.J. Yates & A. D. Miller (Eds.), *Vestibular autonomic regulation* (pp. 147-173). Boca Raton, FL: CRC Press.

Moshell, J., Blau, B. Knerr, B., Lampton, D., & Bliss, J. (1993). A research testbed for virtual environment training applications. *Proceedings of the IEEE Virtual Reality Annual International Symposium* (pp. 83-89). New York: Institute of Electrical and Electronics Engineers, Inc.

Moshell, J. M., & Hughes, C. E. (2002). Virtual environments as a tool for academic learning. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 893-910). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Munro, A., Breaux, R., Patrey, J., & Sheldon, B. (2002). Cognitive aspects of virtual environment design. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 415-434). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Naval Training Systems Center (NTSC). (1989, October). *Simulator sickness field manual: MOD 4*. Orlando, FL: Naval Training Systems Center, Human Factors Laboratory.

Neale, H. R., Brown, D. J., Cobb, S. V. G., & Wilson, J. R. (1999). Structured evaluation of virtual environments for special-needs education. *Presence, 8*(3), 264-282.

Nelson, W. T., Bolia, R. S., Roe, M. M., & Morley, R. M. (2000). Assessing simulator sickness in a see-through HMD: Effects of time delay, time on task, and task complexity. *Proceedings of the 2000 IMAGE Conference* (pp. 250-256).

North, M. M., North, S. M., & Coble, J. R. (2002). Virtual reality therapy: An effective treatment for psychological disorders. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 1065-1078). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Padmos, P., & Milders, M. (1992). Quality criteria for simulator images: A literature review. *Human Factors, 34*(6), 727-748.

Park, A. H., & Hu, S. (1999). Gender differences in motion sickness history and susceptibility to optokinetic rotation-induced motion sickness. *Aviation, Space, and Environmental Medicine, 70*(11), 1077-1080.

Pausch, R., Crea, T., & Conway, M. (1992). A literature survey for virtual environments: Military flight visual systems and simulator sickness. *Presence, 1*(3), 344-363.

Pepper, R. L., Smith, D. C., & Cole, R. E. (1981). Stereo TV improves operator performance under degraded visibility conditions. *Optics Engineering, 20*(4), 579-585.

Reason, J. T. (1969). Motion sickness – Some theoretical considerations. *International Journal of Man-Machine Studies, 1*, 21-38.

Reason, J. T. (1978). Motion sickness adaptation: A neural mismatch model. *Journal of the Royal Society of Medicine, 71*, 819-829.

Reason, J. T., & Brand, J. J. (1975). *Motion sickness*. New York: Academic Press.

Regan, E. C., & Price, K. R. (1994). The frequency of occurrence and severity of side-effects of immersion virtual reality. *Aviation, Space, and Environmental Medicine, 65*, 527-530.

Regian, J. W., & Shebilske, W. L. (1992). Virtual reality: An instructional medium for visual-spatial tasks. *Journal of Communication, 42*(2), 136-149.

Reschke, M. F. (1990). Statistical prediction of space motion sickness. In G. H. Crampton (Ed.), *Motion and space sickness* (pp. 263-316). Boca Raton, FL: CRC Press, Inc.

Riccio, G. E., & Stoffregen, T. A. (1991). An ecological theory of motion sickness and postural instability. *Ecological Psychology, 3*(3), 195-240.

Rinalducci, E. J. (1996). Characteristics of visual fidelity in the virtual environment. *Presence, 5*(3), 330-341.

Riva, G., Botella, C., Légeron, P., & Optale, G. (Eds.). (2004). *Cybertherapy: Internet and virtual reality as assessment and rehabilitation tools for clinical psychology and neuroscience*. Amsterdam: IOS Press.

Riva, G., Wiederhold, B., & Molinari, E. (Eds.). (2000). *Virtual environments in clinical psychology and neuroscience*. Amsterdam: IOS Press.

Rizzo, A. A., Buckwalter, J. G., & van der Zaag, C. (2002). Virtual environment applications in clinical neuropsychology. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 1027-1064). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rushton, S., Mon-Williams, M., & Wann, J. P. (1994). Binocular vision in a bi-ocular world: New generation head-mounted displays avoid causing visual deficits. *Displays, 15*(4), 255-260.

Salas, E., Oser, R. L., Cannon-Bowers, J. A., & Daskarolis-Kring, E. (2002). Team training in virtual environments: An event-based approach. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 873-892). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Satava, R. M., & Jones, S. B. (2002). Medical applications of virtual environments. In L. J. Hettinger & M. Haas (Eds.), *Virtual and adaptive environments: Applications, implications, and human performance issues* (pp. 325-343). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Satava, R. M., & Jones, S. (2003). Medical applications of virtual reality. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 937-957). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Salzman, M. C., Dede, C., & Loftin, R. B. (1995). Usability and learning in educational virtual realities. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 486-490). Santa Monica, CA: Human Factors and Ergonomics Society.

Servignat, C., Flores, J., Kemeny, A., & Vernet, M. (1995). The role of a driving simulator in an ergonomic evaluation procedure: The case of an on-board aid system. *Proceedings of the Driving Simulation Conference "DSC95"* (pp. 77-93). Montigny-le-Bretonneux: Neuf Associes.

Shewchuk, J. P., Chung, K. H., & Williges, R. C. (2002). Virtual environments in manufacturing. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 1119-1141). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Smart, L. J., Jr., Stoffregen, T. A., & Bardy, B. G. (2002). Visually induced motion sickness predicted by postural instability. *Human Factors, 44*(3), 451-465.

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 42nd Annual Meeting* (pp. 1476-1480). Santa Monica, CA: Human Factors and Ergonomics Society.

Stanney, K. M., Kennedy, R. S., & Drexler, J. M. (1997). Cybersickness is not simulator sickness. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 1138-1142). Santa Monica, CA: Human Factors and Ergonomics Society.

Stanney, K. M., Kennedy, R. S., & Kingdon, K. (2002). Virtual environment usage protocols. In K.M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 721-730). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Stanney, K. M., Kingdon, K. S., & Kennedy, R. S. (2002). Dropouts and aftereffects: Examining general accessibility to virtual environment technology. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 2114-218). Santa Monica, CA: Human Factors and Ergonomics Society.

Stanney, K. M., Lanham, D. S., Kennedy, R. S., & Breaux, R. (1999). Virtual environment exposure drop-out thresholds. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 1223-1227). Santa Monica, CA: Human Factors and Ergonomics Society.

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.

Stanney, K., Salvendy, G., Deisinger, J., DiZio, P., Ellis, S., Ellison, J., Fogleman, G., Gallimore, J., Singer, M., Hettinger, L., Kennedy, R., Lackner, J., Lawson, B., Maida, J., Mead, A., Mon-Williams, M., Newman, D., Piantanida, T., Reeves, L., Riedel, O., Stoffregen, T., Wann, J., Welch, R., Wilson, J., & Witmer, B. (1998). Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *International Journal of Human-Computer Interaction, 10*(2), 135-187.

Stern, R. M., Hu, S., Vasey, M. W., & Koch, K. L. (1989). Adaptation to vection-induced symptoms of motion sickness. *Aviation, Space, and Environmental Medicine, 60*(6), 566-572.

Stern, R. M., Hu, S., Anderson, R. B., Leibowitz, H. W., & Koch, K. L. (1990). The effects of fixation and restricted visual field on vection-induced motion sickness. *Aviation, Space, and Environmental Medicine, 61*(8), 712-715.

Stern, R. M., Koch, K. L., Leibowitz, H. W., Lindblad, I. M., Shupert, C. L., & Stewart, W. R. (1985). Tachygastria and motion sickness. *Aviation, Space, and Environmental Medicine, 56*(11), 1074-1077.

Stoffregen, T. A., Hettinger, L. J., Haas, M. W., Roe, M. M., & Smart, L. J. (2000). Postural instability and motion sickness in a fixed-base flight simulator. *Human Factors, 42*(3), 458-469.

Stoffregen, T. A., & Riccio, G. E. (1991). An ecological critique of the sensory conflict theory of motion sickness. *Ecological Psychology, 3*(3), 159-194.

Stoffregen, T. A., & Smart, L. J., Jr. (1998). Postural instability precedes motion sickness. *Brain Research Bulletin, 47*(5), 437-448.

Stone, R. J. (2002). Applications of virtual environments: An overview. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 827-856). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Stott, J. R. R. (1990). Adaptation to nauseogenic motion stimuli and its application in the treatment of airsickness. In G. H. Crampton (Ed.), *Motion and space sickness* (pp. 373-390). Boca Raton, FL: CRC Press, Inc.

Treisman, M. (1977). Motion sickness: An evolutionary hypothesis. *Science, 197*, 493-495.

Triggs, T. J., & Fronsko, A. (1995). Scenarios, databases, and performance measures for simulator based driver training curricula. *Proceedings of the Driving Simulation Conference "DSC95"* (pp. 395-408).

Turnage, J. J., Kennedy, R. S., Gilson, R. D., Bliss, J. P., & Nolan, M. D. (1988). *The use of surrogate measurement for the prediction of flight training performances*. Orlando, FL: University of Central Florida, Institute for Simulation and Training.

Turnage, J. J., & Lane, N. E. (1987). The use of surrogate techniques for the measurement of team performance. *Proceedings of the 31st Annual Meeting of the Human Factors Society* (pp. 638-642). Santa Monica, CA: Human Factors and Ergonomics Society.

Uliano, K. C., Kennedy, R. S., & Lambert, E. Y. (1986). Asynchronous visual delays and the development of simulator sickness. *Proceedings of the Human Factors Society 30th Annual Meeting* (pp. 422-426). Santa Monica, CA: Human Factors and Ergonomics Society.

Uliano, K. C., Lambert, E. Y., Kennedy, R. S., & Sheppard, D. J. (1986). *The effects of asynchronous visual delays on simulator flight performance and the development of simulator sickness symptomatology* (Final Report No. NAVTRASYSCEN 85-C-0024-1). Orlando, FL: Essex Corporation.

Ungs, T. J. (1987). Simulator induced syndrome: Evidence for long term simulator aftereffects. *Proceedings of the Human Factors Society 31st Annual Meeting* (pp. 505-509). Santa Monica, CA: Human Factors and Ergonomics Society.

Ungs, T. J. (1988). Simulator induced syndrome in Coast Guard aviators. *Aviation, Space, and Environmental Medicine, 59*(3), 267-272.

Van Hoy, B. W., Allgood, G. O., Lilienthal, M. G., Kennedy, R. S., & Hooper, J. M. (1987). Inertial and control systems measurements of two motion-based flight simulators for evaluation of the incidence of simulator sickness. *Proceedings of the IMAGE IV Conference* (pp. 265-273). Phoenix, AZ: Image Society Incorporated.

Viirre, E., & Ellisman, M. (2003). Vertigo in virtual reality with haptics: Case report. *Cyberpsychology and Behavior, 6*(4), 429-431.

Wann, J. P., & Mon-Williams, M. (2002). Measurement of visual aftereffects following virtual environment exposure. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 731-749). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Warner, H. D., Serfoss, G. L., Baruch, T. M., & Hubbard, D. C. (1993). *Flight simulator-induced sickness and visual displays evaluation* (Final Technical Report No. AL/HR-TR-1993-0056). Brooks AFB: Armstrong Laboratory, Air Force Material Command.

Warwick-Evans, L., & Beaumont, S. (1995). An experimental evaluation of sensory conflict versus postural control theories of motion sickness. *Ecological Psychology, 7*(3), 163-179.

Warwick-Evans, L. A., Church, R. E., Hancock, C., Jochim, D., Morris, P. H., & Ward, F. (1987). Electrodermal activity as an index of motion sickness. *Aviation, Space, and Environmental Medicine, 58*, 417-423.

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.

Welch, R. B. (2002). Adapting to virtual environments. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 619-636). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wertheim, A. H. (1999, November). *The assessment of aftereffects of real and simulated self motion: Motion sickness and other symptoms* (TNO Report No. TM-99-A074). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Wilson, J. R. (1996). Effects of participating in virtual environments: A review of current knowledge. *Safety Science, 23*(1), 39-51.

Wilson, J. R. (1997). Virtual environments and ergonomics: Needs and opportunities. *Ergonomics, 40*(10), 1057-1077.

Wilson, J. R. (1999). Virtual environment applications and applied ergonomics. *Applied Ergonomics, 30*(1), 3-9.

Witmer, B. G., Bailey, J. H., & Knerr, B. W. (1996). Virtual spaces and real world places: Transfer of route knowledge. *International Journal of Human-Computer Studies, 45*, 413-428.

Wright, R. H. (1995, June). *Helicopter simulator sickness: A state-of-the-art review of its incidence, causes, and treatment* (ARI Research Report No. 1680). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Yardley, L. (1992). Motion sickness and perception: A reappraisal of the sensory conflict approach. *British Journal of Psychology, 83*, 449-471.

Youngblut, C. (1998, January). *Educational uses of virtual reality technology* (IDA Technical Report No. D-2128). Alexandria, VA: Institute for Defense Analyses. Available: http://www.hitl.washington.edu/scivw/youngblut-edvr/D2128.pdf