6-28-2022

Validation of the GUESS-18 for the Usability of a Virtual Reality Racing Game

William A. Carman
Louisiana State University and Agricultural and Mechanical College

VALIDATION OF THE GUESS-18 FOR THE USABILITY OF A VIRTUAL REALITY RACING GAME

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Mechanical and Industrial Engineering

by
William Carman
B.S., University of Montevallo, 2018
M.S., Louisiana State University, 2021
August 2022

Table of Contents

Abstract iv

Introduction 1

Literature Review 3
- Usability Heuristics 3
- Questionnaires for Usability Testing 4
- Applications of Virtual Reality 5
- The GUESS-18 7
- Survey Validation 8
- Purpose of Study 10

Project Proposal 11
- Experimental Design 11
- Participants 11
- Setting 11
- Dependent Variables 16
- Independent Variable 17
- Procedures 17
- Data Analysis 18

Results 20
- Performance 20
- Usability Scores 21
- Confirmatory Factor Analysis for the GUESS-18 in VR 24

Discussion 27
- Usability 27
- Simulator Sickness 28
- Validation 29
- Relevance to Modern Research 30
- Limitations and Future Research 31

Conclusion 33

Appendix A. GUESS-18 Questionnaire 34

Appendix B. SUS Questionnaire 35

Appendix C. Simulator Sickness Questionnaire 36

Appendix D. IRB Forms and Survey Links 37
- IRB Consent Form 37
- IRB Approval 39
- IRB Amended Approval 41
- Survey Links 43

References 44

Vita 51

Abstract

Virtual reality (VR) technology has been in development for many decades, with recent strides in the consumer market. Head-mounted displays (HMDs) provide an immersive VR experience by inserting users into an artificially constructed digital world. However, there is currently no standardized tool for measuring the usability of a VR system or environment. The GUESS-18 is a validated usability questionnaire designed for measuring the usability of video games with 9 factors present in video games. The objective of this study was to validate the GUESS-18 for measuring the usability of VR in a gaming environment.

Participants played Assetto Corsa, a racing game, with an HMD in VR and with a traditional monitor. The best lap times were recorded as a performance measure. Errors were also recorded, which included driving off the track or colliding with a wall. Users answered the GUESS-18 and the System Usability Scale (SUS) after their experience. They also answered the Simulator Sickness Questionnaire (SSQ) before and after using VR. Confirmatory Factor Analysis (CFA) was performed to determine the validity of the GUESS-18's measurements for usability of the racing game in VR.

The GUESS-18 is an accurate and effective tool for measuring usability of the game in VR. The Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA) values indicate appropriate model fit, with values of 0.967 and 0.056, respectively. VR yielded significantly faster lap times by about 5%. No other differences were found for performance. Usability scores from the GUESS-18 were significantly higher in the VR condition than the traditional monitor by about 6.4%. SSQ scores were significantly higher after engaging with VR by 282.5%. No differences were found for SUS scores between display conditions. We also found that playing the game in VR resulted in greater immersion and personal gratification than playing with the monitor. Future research should focus on the development and validation of a generalized VR usability tool that captures the latent factors when using a VR system.

Introduction

Usability is one of the most important elements when working with user-centered products. From vehicles to hand tools, usability provides satisfying experiences through appropriate implementation of usability heuristics. The importance of usability has given rise to surveys designed for measuring perceived usability. These surveys are broad in scope for general applicability or narrow for specific domains. Despite the variety of usability tools available, there is no tool designed specifically for measuring perceived usability of Virtual Reality (VR) applications. This gap motivates the present study through an implementation of a validated usability survey centered around video games. This survey is the Game User Experience Satisfaction Scale (GUESS-18), a psychometrically validated survey designed to measure perceived usability of video games. The objective of this study is to validate the GUESS-18 for perceived usability of a VR video game.

VR is generally perceived as emerging technology (Fang & Lin, 2019), with expected advancements in the coming years. It provides an artificial environment in which users can manipulate or interact with their surroundings digitally. This digital world can take on different settings and conditions to replicate virtually any situation, location, or condition. A VR environment can take form as a game or working office; It can be a training simulation or educational experience. The malleability of these digital worlds provides opportunities that were previously inaccessible by physical limitations. VR has made its way into the mainstream with intentions of utilizing VR technology to connect users through a virtual world, merging reality with VR. With the increasing presence of VR, we must study its applications to fully utilize its benefits.

VR is an important factor in the video game industry. The gaming industry has steadily grown over the past decade, with VR as one of the most prominent technologies in the field. There are a wide variety of VR applications in the realm of video games. For instance, Beat Saber is a popular VR game that has sold over 4 million copies since its release in 2018 through February of 2021. It is estimated that the game has generated over $100 million in revenue within that time (Baker, 2021). Other games have been developed without VR in mind, but growth from VR motivated developers to retroactively incorporate VR features into their games, like with Assetto Corsa. VR and video games share many characteristics. Both involve digitally created environments that are manipulable by the user. Interaction with the digital world stimulates user engagement, leading to a fulfilling user experience. As VR gaming expands, the need for VR-centered usability tools expands as well.

VR has taken on a wide variety of beneficial manifestations besides that of video games, such as through education and training. VR is regularly used in military procedures, flight training amongst pilots, and even in the medical field. Safety is a major advantage when training in VR since there is minimal risk to the user. VR also allows for easily repeatable scenarios, reducing waste from consuming materials that come with physical simulation training.

There are a wide variety of tools used to measure perceived usability. Tools such as the System Usability Scale (SUS) or the Questionnaire for User Satisfaction (QUIS) are commonly used when measuring general usability (Brooke, 1986) (Chin et al., 1988). These tools have tested usability in a wide variety of systems. The 18-item Game User Experience Satisfaction Scale-18 (GUESS-18) was recently developed to measure the usability of video games using nine subscales common in video games (Keebler et al., 2020). While other tools exist for measuring video game usability, the GUESS-18 is specific in its subscales and has been psychometrically validated with exploratory and confirmatory factor analysis. The objective of this study is to validate the GUESS-18 for usability in VR using a specific game, Assetto Corsa.

Literature Review

The purpose of this literature review is to discuss modern work in the topics of usability and applications within the setting of VR. By understanding this background, we identify the gap in research: the need for a validated survey centered around perceived usability of VR. Our attempt to explore this gap involves the implementation of the GUESS-18 and the SUS as a benchmark for usability testing. While neither of these tools are specific to VR, we intend to detect if they accurately capture the perceived usability of a VR video game.

Usability Heuristics

Jakob Nielson from the Nielson Normal Group established 10 heuristics for general interface design in 1994 (Nielsen, 1994). These heuristics are widely accepted today as the gold standard for usability in interface design. They provide guidelines that impact the overall usability of an interface. Note that VR itself is a special kind of interface, so the heuristics are still relevant for VR usability. Each heuristic is listed below.

The visibility of a system's status should always keep users informed about what is going on with the system. The interface should clearly communicate the state of the system, as well as important changes that occur. Understanding a system's status is important for decision making and improved usability.

There should be a matching between the system and the real world. Effective designs use language, icons, and concepts that are familiar to the user to develop a natural mapping of the system. It is easier for users to learn interfaces that incorporate these practices due to familiarity.

Users should have adequate control and freedom over the interface, especially when undoing mistakes; They might find themselves in an unwanted situation. In moments like this, good designs should provide an emergency exit or an opportunity to undo decisions easily.

The system should stay consistent with the language and visuals of outside systems. Users naturally develop preconceived ideas about how an interface should look due to past experiences with similar systems. When designing an interface, developers should consider adopting designs from similar applications to align with user expectations.

Error prevention is essential for facilitating usable designs. Errors can be deferred with messages or indicators, but optimal designs limit the opportunity for mistakes in the first place. The two types of errors are slips and mistakes. Slips are unintentional and unconscious while mistakes result from a conceptual misunderstanding of how the system truly works. Interfaces with high usability are designed to minimize both types of errors.

When retrieving information from memory, individuals can either recognize or recall that information. Recognition means that users identify elements that they are familiar with

to bring the information to memory, rather than directly recalling memorized information. Generally, recognizing is easier than recalling information for humans. Efficient interface designs apply recognition techniques by showing users the information that help them make the best decisions.

Flexibility allows for multiple ways to complete a task. This occurs frequently in the form of shortcuts, which speed up task operations for experienced users. A flexible interface accommodates users of all skill levels.

An interface's aesthetic plays a significant role in its usability. The design should only incorporate visuals that are relevant to the needs of the user. A popular choice of aesthetic uses minimalist designs by limiting information since too much clutter inhibits user flow.

The interface should identify errors and notify the user when they occur. Errors should be clearly marked, and messages should intuitively explain what has gone wrong. Corrective action should also be suggested to solve the issue.

Help must be provided in an easily identifiable manner. Optimal systems should require no additional explanation to operate but providing help can assist those who need further guidance. Help should be easily identifiable and structured in a digestible way. It should be intuitively written and incorporate a minimalist design.

Questionnaires for Usability Testing

Since usability is an integral component of the discussion, the definition of a usable system must be understood. The term "usability" generally refers to how easily a user can interact with an interface, product, or other user-centered medium. It is defined by the effectiveness, efficiency, satisfaction, and frequency of errors in a system. Usability is concerned with factors such as learnability, ease-of-use, user-satisfaction, intuitiveness, and system inconsistencies. Usability testing measures common patterns and psychological conditions demonstrated by human beings. The most common measures for assessing usability stem from the definition of usability itself: success rate for task performance, time taken when performing a task, frequency of errors, and user satisfaction (Nielsen, 1994). Since user satisfaction is impacted by performance, it is sufficient to measure the user's subjective experience to quantify usability. User satisfaction can be measured with questionnaires that obtain user feedback. This feedback can be informal or attained from questionnaires administered after participants interact with the product. A 5 or 7-point Likert scale is commonly used for scoring these questionnaires.

The System Usability Scale (SUS) is one of the most prevalent tools for measuring perceived usability (Brooke, 1986). The SUS is composed of 10 items, alternating between positive and negative toned questions measured on a 5-point Likert scale. The SUS was used in 43% of post-study questionnaires in industrial usability studies by 2009 (Lewis & Sauro, 2009). It is widely accepted as the standardized questionnaire for perceived usability (Lewis, 2018). Since its release in 1986 (Brooke, 1986), the SUS

has been used across numerous topics for usability evaluation. For example, the SUS was used to evaluate perceived usability of Microsoft Teams as a learning medium during the COVID-19 pandemic, with results showing that users were likely to recommend the program to others (Pal & Vanijja, 2020). Liang and colleagues used SUS scores to evaluate usability of mainstream fitness devices (2018). They suggested that perceived usability of these devices was unsatisfactory and in need of improvement. These distinct topics highlight the diversity of systems that the SUS can measure. Flexibility is a driving factor for why the SUS is used in the present study.

Other than the SUS, there are a range of other usability tools still used today. The Questionnaire for User Interaction Satisfaction (QUIS) is a 27-item survey developed to measure usability of human-computer interfaces. It was found to possess strong reliability across a variety of interfaces (Naeini et al., 2015). Another well-used, reliable tool for testing usability is the Software Usability Measurement Inventory (SUMI). This 50-item survey provides a method of comparison between different versions of the same item, providing objective assessments of user satisfaction with the system (Arh & Blažič, 2008). The Post Study System Usability Questionnaire (PSSUQ) is another established tool designed to measure users' perceived satisfaction with computer systems. It consists of 19-items and has been found to possess validity, reliability, and sensitivity (Lewis, 2002). While these questionnaires are recognized as effective testing methods, they lack the generality and brevity that the SUS provides.

VR usability is traditionally studied with usability surveys. The Game Experience Questionnaire (GEQ) was used to determine relationships between VR game displays and traditional 2D monitor displays (Pallavicini & Pepe, 2019). They found game performance was mostly unchanged between these two mediums, but player engrossment, flow, and intense positive emotions were stronger when playing in VR. Another study found similar results when using the Visual Analog Scale (VAS) with anxiety (VAS-A), happiness (VAS-HP), and surprise (VAS-SP) (Pallavicini et al., 2019). This study also showed players exhibited more intense emotional responses when playing games in VR. Research continues to advance to better our understanding of VR satisfaction. The tools previously discussed measure usability for specific items like video games or computer interfaces, but VR uses newer technology (Head-mounted displays (HMD), VR interfaces) that may not be accurately measured with previous usability tools. As VR advances, the need for a validated tool centered around VR usability advances as well.

Applications of Virtual Reality

In this context of the present work, VR refers to the use of an HMD to immerse the user into an artificially constructed, digital world. However, other forms of digital environments exist beyond the realm of VR. Augmented Reality (AR) overlays digital elements on the physical world to enhance the perceived environment. AR allows the virtual world and physical world to be experienced simultaneously, with the human interacting with the physical world and observing elements from the digital world (Mann et al., 2018). AR differs from VR in the sense that both the physical and digital worlds

are experienced with AR, but VR fully immerses the user in a digital environment. Another example of an alternative reality is Mixed Reality (MR). MR is a blend of VR and AR, allowing the user to interact with both digital and physical worlds at the same time (Mann et al., 2018). MR provides a level of interaction beyond that of AR since AR allows users to perceive the digital world, but not interact with it. Each of these realities enhances the human's perception of the world by providing additional information and interaction outlets.

Extensive work was conducted for applications of VR-centered training. VR training refers to using virtual simulations to mimic real-world scenarios for the development of real-world skills. The impacts of VR training can be attributed to gamification. Gamification is the application of game elements and terminology to non-game environments. There is evidence that supports gamification as an effective means for education (Zainuddin et al., 2020). A virtual simulation is the gamification of an environment, highlighting the potential of VR for developmental skills.

Simulation training is beneficial for improving medical performance of staff and patient morbidity and mortality (Martin et al., 2020). First responders have used VR in their training with positive user feedback on the experience (Mossel et al., 2021). VR training even expands to the realm of aircraft cabin safety procedures (Buttussi & Chittaro, 2018). Their study focused on display types and found significant differences between a 2D monitor and VR headset for user engagement and presence. There is evidence suggesting VR elicits empathy and engagement more so than traditional 2D displays (Schutte & Stilinović, 2017).

VR provides a safe space in which users can train without exposure to real-world hazards. Chemical, Biological, Radiological, and Nuclear crisis training implements VR to minimize exposure with severe hazards (Mossel et al., 2017). Khooshabeh applied VR to a tank commander to coordinate virtual tank positions using the TALK-ON communication system (Khooshabeh et al., 2017). VR was found to be effective for reducing burn-induced pain and managing other pain with cognitive behavioral therapy techniques (Sharar et al., 2008). Spatial knowledge obtained through VR training has shown to transfer to real-world scenarios as effectively as real-world training when exposed for an extended period of time (Waller et al., 1998).

VR has vast beneficial applications in the medical field. VR has been used to improve surgical techniques (Javaid & Haleem, 2020). It also has been applied to neurology and cardiology for monitoring patients' outcomes (Javaid & Haleem, 2020). VR employs 3D technology to further understanding of human anatomy for treatment or education (Haleem & Javaid, 2018). VR-based rehabilitation was found to be a viable method for improving patients' balance and gait (Porras et al., 2019). VR has been proposed to assist with healthcare of COVID-19 through its educational outlets and relevant resources (Singh et al., 2020). These are only a small portion of recent studies that indicate VR benefits in the medical field.

VR finds its utility in various fields when performing in serious games. Serious games are defined as games that engage the player through means beyond simple entertainment (Susi et al., 2007). They allow users to interact with digital scenarios that function to facilitate learning (Bente & Breuer, 2010). Fox and colleagues applied a

serious entrepreneurship game in which users were given a business to conduct (2018). Their research indicated that their game played a part in the learning process. The game allowed users to experience running a business which led to educational growth (Fox et al., 2018).

Proper training often requires access to equipment or materials that are costly to obtain. These materials may be consumable, leading to repeated purchases and regular expenses. Purchasing a VR headset could prove to be more cost effective in the long-term by mitigating these regular expenses. Furthermore, training may require individuals to operate at a specific location. A VR headset allows training to occur almost anywhere, minimizing travel time. As technology develops, virtual simulations provide opportunities in more cost-effective methods compared to conventional training.

The GUESS-18

The GUESS-18 used in the current study is based on the Game User Experience Satisfaction Scale (GUESS), a 55-item survey designed to measure the usability of video games (Phan et al., 2016). The GUESS uses nine subscales to fully capture video game usability. These subscales are playability, narratives, play engrossment, enjoyment, creative freedom, audio aesthetics, personal gratification, social connectivity, and visual aesthetics. Each of these factors is a primary component for the user-friendliness of video games. The term “playability” is an alternative to usability within the context of a game. This tool was psychometrically validated to possess content validity, internal consistency, and both convergent and divergent validity (Phan et al., 2016). The GUESS was tested by assessing over 450 unique games across popular genres with the implication that it applies to a variety of different game types (Phan et al., 2016).

The GUESS was used to measure usability in serious games. Fussel and Hight used the GUESS for VR flight training usability testing. They compared the flight training effectiveness between 2D displays and VR with results indicating a significant difference in usability between the groups; VR was found to have improved usability (Fussell & Hight, 2021). Another team cited the GUESS as a critical resource for evaluating video game usability for prosthetic arm training (Manero et al., 2018). Fussel and colleagues used a modified version of the GUESS for a flight VR tutorial (2019). This shortened version of the GUESS measured usability, immersion, playability, enjoyment, personal gratification, and visual aesthetics (Fussell et al., 2019). The GUESS was used for a game that taught fourth graders about health education through an interactive medium (Yoshimura, 2021); Results indicated a lack of user learnability from the game. The GUESS measured usability for a back-extension glide VR game involving patients with chronic back pain. The apparatus was deemed to have moderate to high usability for patients with moderate chronic back pain. (Bateye et al., 2020).

Despite the GUESS’s success, a 55-question survey is cumbersome when administering frequent assessments; The GUESS takes about 10-15 minutes to complete (Keebler et al., 2020). Because of this, the GUESS-18 was derived to provide a more concise testing experience. The GUESS-18 is an 18-item survey that retains the

same nine subscales from the GUESS, but with two specific questions from each of the nine subscales (Keebler et al., 2020). The GUESS-18 was validated with confirmatory factor analysis, as well as convergent and divergent validation methods (Keebler et al., 2020). Results from the analysis indicated that the GUESS-18 delivers a strong measurement system for perceived usability of video games (Keebler et al., 2020).

Since its release in 2019, the GUESS-18 has appeared in several research studies. Schorer and Protopsaltis employed the GUESS-18 for the usability of NEMESIS, an educational game that teaches players about community changes and their consequences (Schorer & Protopsaltis, 2021). GUESS-18 scores indicated that NEMESIS was in the range of neutral to positive in terms of usability. The GUESS-18 measured feasibility of interpersonal emotional regulation of adolescents through a serious adventure game (Mittmann et al., 2021). Furthermore, the GUESS-18 was evaluated to determine its effectiveness of measuring video game satisfaction for gamers with disabilities, with results indicating that a new scale should be made to accommodate gamers with disabilities (Van Ommen & Chaparro, 2021). Another study implemented mixed reality for the popular mobile game Angry Birds. Scores from the GUESS-18 determined significant differences of usability when playing in mixed reality; mixed reality was more engaging than its mobile counterpart (Sinlapanuntakul et al., 2020).

It is worth noting that another questionnaire, the GUESS-24, was introduced in an unpublished manuscript between the GUESS and the GUESS-18. Shelstad et al conducted a study about how user experience scales predict purchasing intent of video games with the GUESS-24, ENJOY, and UEQ-S. Results indicated that the GUESS-24 was the most accurate in predicting purchasing intent of video games (Shelstad et al., 2020). Another study analyzed the relationships between these three tools. The GUESS-24 was deemed to provide the most detail with its nine subscales (Shelstad et al., 2019).

While the GUESS and GUESS-18 were developed for the purpose of evaluating video games, these tools have been expanded to other domains. VR environments share several characteristics with video games such as visuals, engrossment, enjoyment, and playability. Due to the overlapping nature between video games and VR, the GUESS questionnaires are adequate tools for evaluating usability of VR. Since no tool has been designed specifically for measuring VR usability, we validated the GUESS-18 for measuring perceived usability of a VR video game.

Survey Validation

The purpose of a survey is to obtain subjective data of a system's underlying factors. Survey questions aim to measure these factors with users' responses. However, certain questions may not measure the factors that they intend to measure. Surveys must be validated with construct validity to verify that they measure what they claim to measure. Construct validity identifies if a relationship exists between observed variables and latent constructs. Factor Analysis (FA) is the general method for identifying the underlying relationships between variables (Mahmoud & Kamel, 2010). It explains the

covariance between sets of observed data and the latent factors of the model. The techniques of FA are Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA).

EFA is a fundamental tool in the survey validation process. It discovers the underlying factor structure of a system and establishes the baseline model (Child, 1990). It provides the number of factors within the model and is used when there is no prior knowledge of the factor structure (Watkins, 2018). The most common technique in EFA is Principal Component Analysis (PCA), which identifies and reduces the number of significant factors in a model. EFA was used in the validation process of the GUESS (Phan et al., 2016).

PCA is a factor reduction technique that reduces the dimensions of a data set (Wold et al., 1987). It aims to transform large sets of variables into smaller sets that still capture most of the information described by the larger set. Briefly, PCA finds correlations between variables and identifies the principal components in the model. The goals of PCA include simplification and classification of variables, reduction of data, detection of outliers, and foundational modeling.

Once the basis of a survey's factor structure is established it can be verified with CFA. CFA evaluates how well the model fits observed data. It uses techniques that determine if relationships exist between the observed variables and the underlying factors within the survey (Suhr, 2006). CFA requires a pre-established model for analysis, as opposed to EFA which establishes the model. Techniques such as chi-square, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA) each test model fit. Collectively, they form the backbone of CFA.

The chi-square test provides a measure between the expected and observed covariance matrices (Suhr, 2006). Lower chi-square values point to smaller differences between covariance matrices, indicating better model fit. Chi-square values can be computed by comparing the standardized scores between instruments.

The CFI assesses model fit by reviewing the difference between the observed data and the predetermined model. CFI values range from 0 to 1, with higher values indicating better fit of the model. The threshold of 0.9 and above is generally accepted as an acceptable level of fit (Hu & Bentler, 1999). To obtain CFI values, take the quotient from the observed model from that of the baseline model, then subtract that value from 1.

The RMSEA examines residuals of the model by calculating the standard deviation of residuals. RMSEA values range from 0 to 1, with smaller values indicating a better model fit. Values of 0.06 and below are typically recognized as acceptable (Hu & Bentler, 1999).

There are other types of validity that must be considered when validating a survey. Content validity evaluates if a survey is representative of the metrics it intends to obtain. This is tested of rationale or from feedback with field experts, or relevant literature on the topic. The GUESS was found to possess content validity, supporting the claim that the GUESS-18 has content validity as well (Phan et al., 2016).

Reliability evaluates the internal consistency of the observed data. Reliability is associated with Cronbach's alpha (or coefficient alpha). Cronbach's alpha ranges from 0 to 1, with higher values indicating higher internal consistency. Generally, if test items are highly correlated, then their alpha value will be higher. It is expected that questions within the same subscale of the GUESS-18 will have higher alpha values. Values of 0.7 and above are considered acceptable for this study.

Purpose of Study

This review has covered a wide variety of topics, from interface usability heuristics to VR applications and even survey validation methods. Usability is an essential component of any interface or product, so effective evaluation methods of usability are a necessity in the development process. VR has made great strides in recent years but lacks the targeted usability tools for the developing technology. The success of VR improvement can be facilitated by applying validated usability tools that accurately measure the user experience. While the development of a general usability tool is needed, this research fills the smaller gap of VR gaming usability. This was accomplished through the validation of the GUESS-18 when measuring usability of the video game Assetto Corsa in VR.

Project Proposal

Experimental Design

A randomized within-subjects design was implemented for this experiment. The experiment has 6 dependent variables and 1 independent variable. The independent variable is the display condition: VR and 2D monitor. The dependent variables in this study are the GUESS-18 scores, SUS scores, Simulator Sickness Questionnaire (SSQ) scores, number of minor errors, number of major errors, and best lap time. Questionnaire responses provided the necessary data for validating the GUESS-18. Chi-square, RMSEA, CFI, and Cronbach's alpha were each calculated with survey responses to determine the validity of the GUESS-18. We compared GUESS-18 scores between display conditions to understand how game usability was influenced by display. Performance measures were also compared between display conditions. We hypothesized that the VR condition would stimulate greater user engagement, immersion, and presence than the 2D monitor. GUESS-18 scores and SUS scores were compared for differences in usability scores, with the expectation that no differences would be present. We further hypothesized that performance would be unchanged in the VR condition.

Participants

Participants were LSU students, with a total of 47 participants (28M, 19F) to provide sufficient statistical power. 2 participants did not complete the experiment due to motion sickness during the VR portion of testing; We report results from the remaining 45. The average participant age was 22.5 years (4.2 s.d.) with an average of 5.7 (3.9 s.d.) driving years. Participants were recruited through email and course advertisements to select Industrial Engineering courses as approved by the course instructor, with extra credit as an incentive for participation. All participants had normal or corrected to normal vision and full motor control of their bodies for inclusion. They were at least 18 years of age with no cognitive disabilities that impaired the operation of a motor vehicle. Before testing, each participant was briefed about the testing session and given an opportunity to ask questions. If they agreed to participate, they signed the consent form (see Appendix D) as approved by the Louisiana State University Institutional Review Board IRBAM-22-0217. We also obtained age and gender to categorize the findings of the study and determine any trends between groups.

Setting

Participants played the racing video game Assetto Corsa. This game is a realistic driving simulator that features authentic racing experiences with and without VR. Participants used the Hotlap feature of the game, where they completed laps on a racetrack as fast as possible with no other vehicles present. All participants used the

same vehicle, the Lotus, and drove on the same track. Recordings for each participant's testing session were saved using in-game features if they needed review.

Testing was conducted in the Building Simulation and Information Modeling Construction Management Studio. Participants played Assetto Corsa using two displays: VR and 2D monitor (see figure 1). For the VR setup, participants wore the HTC VIVE Pro Eye HMD (link). This HMD allows for precision eye tracking and high-quality visuals for an immersive experience (see figures 2 and 3).

Figure 1. Layout of testing station

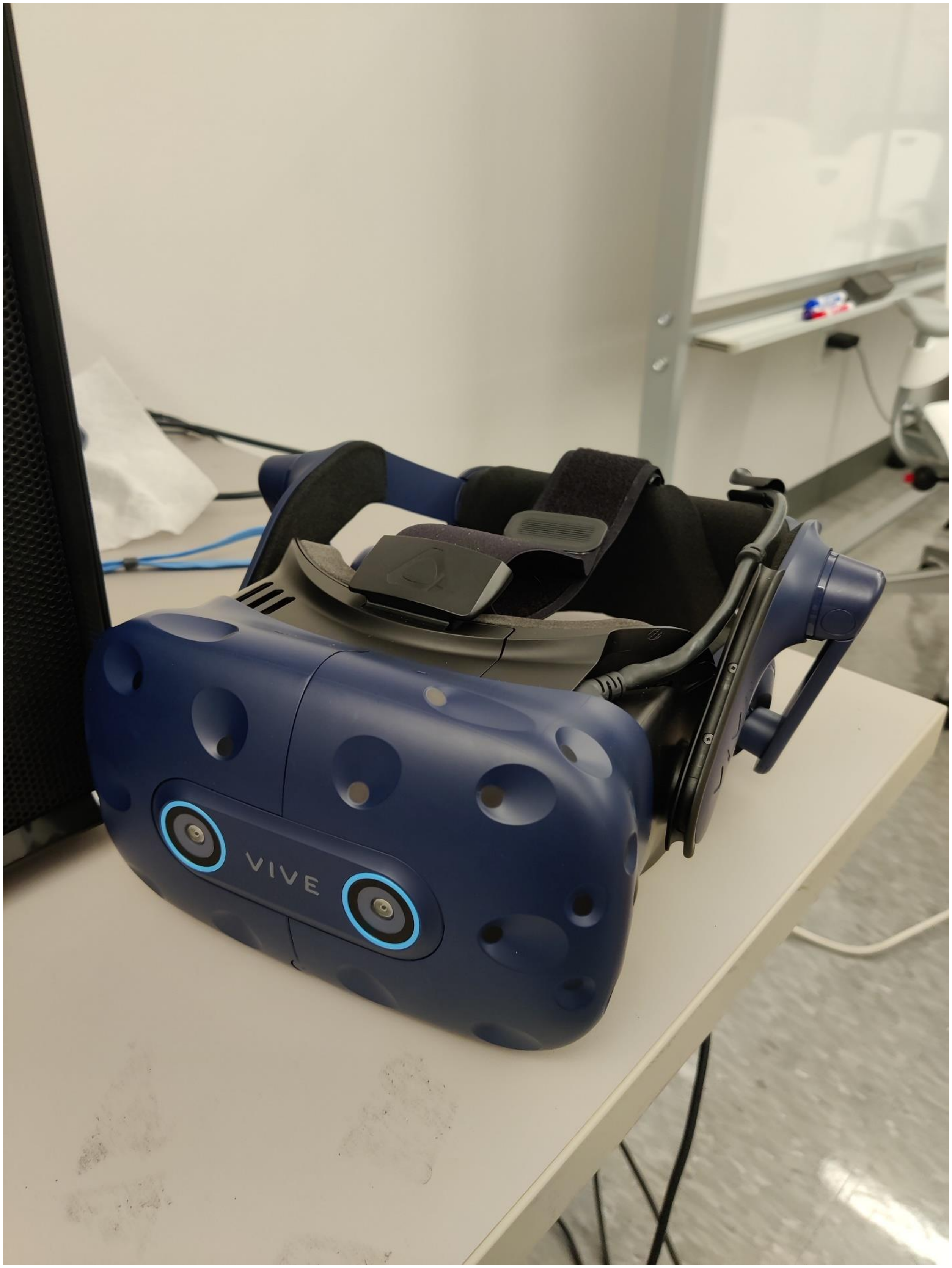

Figure 2. VIVE Pro-Eye HMD from the front

Figure 3. VIVE Pro-Eye HMD from the back

For the 2D display, participants used the BenQ GW2270 21.5” monitor (link). Since this monitor does not have built-in speakers, the AmazonBasics BSK30 external speaker was used to provide sound for the monitor display condition. For both setups, participants will use the Logitech G920 Driving Force Steering Wheel and Floor Pedals as controllers (link).

Dependent Variables

Six dependent variables were measured: GUESS-18 scores, SUS scores, SSQ scores, minor errors, major errors, and best lap time.

Our survey consisted of the GUESS-18 (Keebler et al., 2020) and the SUS (Brooke, 1986) (see Appendices A and B, respectively) for a total of 28 questions. A 7-point Likert scale was used for GUESS-18 evaluation to stay consistent with the work done with the GUESS and GUESS-18. The scoring for the survey was as follows: strongly disagree (1), disagree (2), somewhat disagree (3), neutral (4), somewhat agree (5), agree (6), and strongly agree (7). The GUESS-18 is scored by taking the average values for each subscale, then adding these scores together. Scores range from 9 to 63, with higher scores indicating greater usability. One question must be reverse coded (subtracted from 7) because it is negatively worded (I feel bored while playing the game).

We used a 5-point Likert scale for scoring the SUS. SUS scores are calculated using traditional techniques: Responses are labelled from 1 to 5 (strongly disagree is 1, strongly agree is 5). Odd numbered questions subtract 1 from their score, while even numbered questions subtract their score from 5. These values are added, then multiplied by 2.5 to receive a score out of 100, with higher scores indicating greater usability. An average SUS score is 68, so scores lower than 68 indicate a usability improvement need. All GUESS-18 scores were standardized. This is because these tools have different scaling (GUESS-18 max score is 63 while SUS max score is 100) and must have a common format prior to analysis.

Participants completed the SSQ to monitor participant safety and health (Kennedy et al., 1993) (see Appendix C). The SSQ is composed of 16 items that measure the motion sickness of an individual. The scoring for the survey is as follows: none (0), slight (1), moderate (2), severe (3). A comprehensive score was calculated through the scoring system detailed in its source paper (Kennedy et al., 1993) (see appendix C). Higher scores indicate greater levels of motion sickness. It was administered immediately before engaging with VR to provide a baseline level of motion sickness. After completing testing in VR, participants completed the survey once more to measure how motion sickness levels have changed. This questionnaire was implemented because motion sickness effects are not uncommon when exposed to VR for extended periods of time.

We defined a minor error to be any instance in which a participant drives off the racetrack but does not collide with obstacles off the course. A major error was defined to be any instance where a participant collides with a wall off the racetrack. It is worth

noting that one must first drive off the track to collide with an obstacle, which is why minor errors have the added condition of no collisions. Errors were manually documented during testing sessions. Driving sessions were recorded and reviewed if necessary.

Participants drove for 3 minutes on a track in each display condition. One lap of the track takes around 1:00 to 1:30 to complete, so 2 laps can be completed in 3 minutes. Their fastest lap time was recorded after the session concluded, providing another performance measure.

Independent Variable

The independent variable of this study is the display condition. Participants tested in VR and with a traditional 2D monitor. Testing in both conditions allowed for comparisons of dependent variables to gain insight into performance, usability, and sense of immersion between conditions. This condition also allows for the validation of the GUESS-18 for measuring usability in the context of virtual reality.

Procedures

Participants were tested with two display types: VR headset and 2D monitor. Odd numbered participants started with the VR headset, while even numbered participants will start with the 2D monitor. The SSQ was administered twice, immediately before and after the VR portion of testing. Each participant was given a practice session before testing in which they became familiar with the controls. Practice sessions were held for each display type, totaling 2 practice sessions per participant. After participants completed a driving session, they completed the GUESS-18 and SUS with Qualtrics, a web-based survey tool. Testing concluded after surveys are submitted for both display types. A description of the procedures is provided below and was dependent on participant number:

If the participant number was odd:

1. The experiment was explained to participant and informed consent is obtained.
2. Participant provided demographic information and completed the SSQ.
3. Participant donned the HMD and spent 5 minutes practicing the controls of the game in VR.
4. Participant drove for 3 minutes to complete laps as fast as possible while simultaneously staying on the course.
5. They completed the SSQ, GUESS-18, and SUS in that order.
6. Participant then spent 5 minutes practicing the controls of the game in the monitor condition.
7. Participant drove for 3 minutes to complete laps as fast as possible.
8. Participant completed the GUESS-18 and SUS.
9. Testing concluded.

If the participant number was even:

1. The experiment was explained to participant and informed consent is obtained.
2. Participant provided demographic information.
3. Participant spent 5 minutes practicing the controls of the game with the monitor condition.
4. Participant drove for 3 minutes to complete laps as fast as possible while simultaneously staying on the course.
5. Participant completed the GUESS-18, SUS, and SSQ in that order.
6. Participant donned the HMD and spent 5 minutes practicing the controls of the game in VR.
7. Participant drove for 3 minutes to complete laps as fast as possible
8. They completed the SSQ, GUESS-18, and SUS in that sequence.
9. Testing concluded.

Data Analysis

The preliminary model for latent constructs in the proposed work uses the same nine subscales from the GUESS-18: playability, narratives, play engrossment, enjoyment, creative freedom, audio aesthetics, personal gratification, social connectivity, and visual aesthetics.

To evaluate the effectiveness of a survey, psychometric validation methods was employed. Psychometric methods are commonly used for measuring the quality of standardized questionnaires. The following data was obtained before validating the GUESS-18:

- GUESS-18 scores
- SUS scores
- SSQ scores
- Fastest lap times
- Number of minor errors
- Number of major errors

Once the data was collected, the following metrics (described in the literature review) were obtained to validate the GUESS-18:

- Chi-square (lower values indicate better fit)
- CFI (0.9 and above are accepted)
- RMSEA (0.06 and below are accepted)
- Cronbach's alpha (0.7 and above are accepted)

Final scores from the GUESS-18 and SUS measured overall usability of the system.

We determined statistical differences between display conditions using t-tests with the standard α value of 0.05. Comparisons were between the VR headset and 2D monitor. GUESS-18 and SUS scores were standardized before comparison. Furthermore, we compared differences in driving performance with lap times, as well as the types of

errors from driving. JMP Pro 15, R, and Microsoft Excel were used to conduct data management and analysis.

Results

Performance

The VR condition resulted in significantly faster lap times than the monitor condition ($p = 0.0454$) by about 5% (see figure 4). The average best lap time in the VR condition was 70.2s (9.1 s.d.) while the average best lap time in the monitor condition was 73.8s (11.1 s.d.).

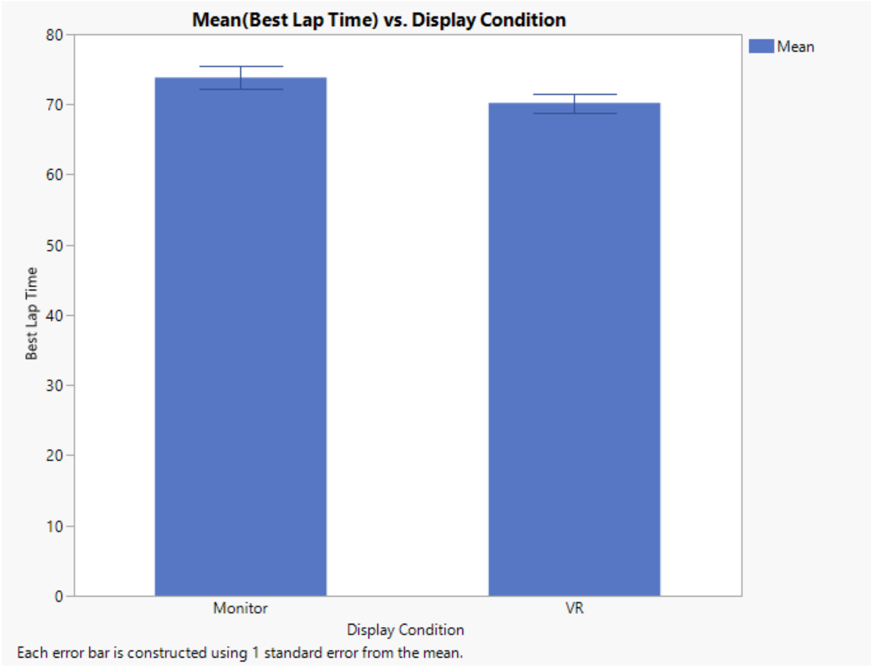

Figure 4. Mean lap times by condition with standard error bar

No significant difference was found for minor errors between display conditions ($p = 0.3115$). The average number of minor errors in VR was 3.3 (3.0 s.d.). The average number of minor errors for the monitor condition was 3.6 (2.5 s.d.). No significant difference was found for major errors between display conditions ($p = 0.2200$). The average number of major errors in VR was 1.9 (2.3 s.d.). The average number of major errors for the monitor condition was 2.3 (2.1. s.d.).

Table 1. Performance measures between display conditions

Display Condition	Mean Lap Time	Mean Minor Errors	Mean Major Errors
VR	70.2s (9.1 s.d.)	3.3 (3.0 s.d.)	1.9 (2.3 s.d.)
Monitor	73.8 (11.1 s.d.)	3.6 (2.5 s.d.)	2.3 (2.1 s.d.)

Usability Scores

GUESS-18 scores were significantly higher in the VR condition than the monitor condition (p = 0.0131) by about 6.4% (see figure 5). The average GUESS-18 score for the VR condition was 50.5 (6.0 s.d.). The average GUESS-18 score for the monitor condition was 47.4 (7.0 s.d.).

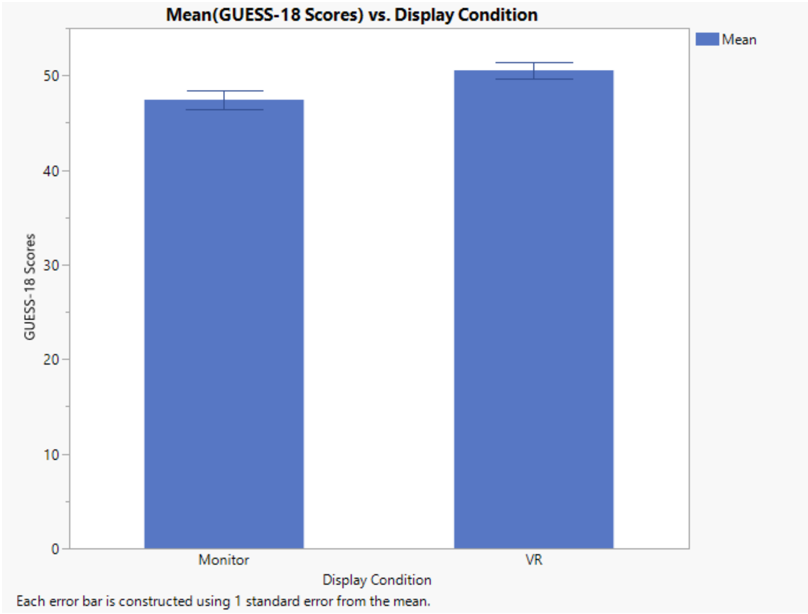

Figure 5. Mean GUESS-18 scores by condition with standard error bar

No significant difference was found for SUS scores between display conditions (p = 0.2839). The average SUS score for the VR condition was 75.4 (13.1 s.d.). The average SUS score for the monitor condition was 76.9 (11.9 s.d.).

SSQ scores were significantly higher after engaging with VR than before engaging with VR (p < 0.0001) (see figure 6). The average SSQ score before VR was 6.3 (11.9 s.d.). The average SSQ score after VR was 24.1 (25.9 s.d.), a 282.5% increase.

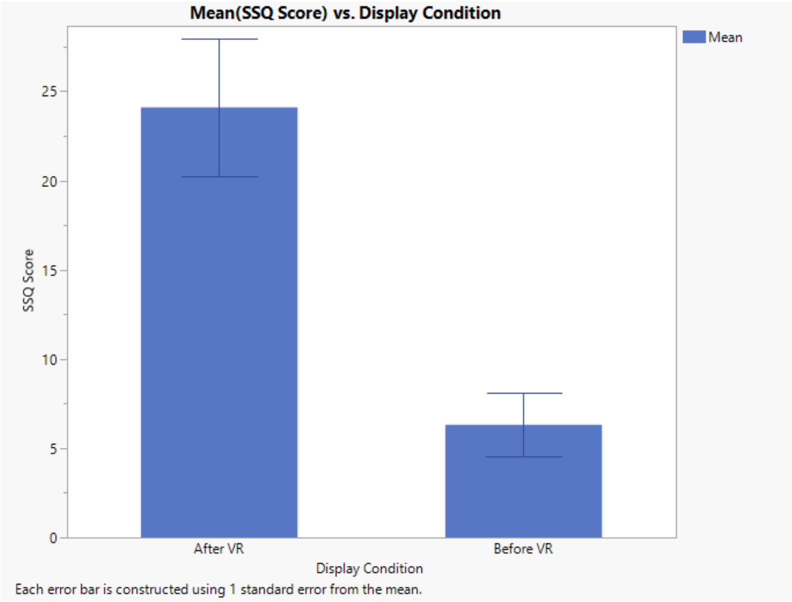

Figure 6. Mean SSQ scores after and before VR with standard error bar

The table below summarize the results of the experiment by display condition.

*Table 2. Dependent measures between display conditions. Variables with a * indicate significant differences.*

Variable	VR Results	Monitor Results
Best Lap Time*	70.2s (9.1s s.d.)	73.8s (11.1s s.d.)
Minor Errors	3.3 (3.0 s.d.)	3.6 (2.5 s.d.)
Major Errors	1.9 (2.3. s.d.)	2.3 (2.1 s.d.)
GUESS-18 Score*	50.5 (6.0 s.d.)	47.4 (7.0 s.d.)
SUS Score	75.4 (13.1 s.d.)	76.9 (11.9 s.d.)
SSQ Score*	6.3 (11.9 s.d.)	24.1 (25.9 s.d.)

We compared standardized survey scores from the GUESS-18 to those from the SUS to determine if there was a difference in usability scores based on display condition. We found that GUESS-18 scores were significantly higher than SUS scores in the VR condition by 6.2% ($p = 0.0243$). The standardized GUESS-18 mean in VR was 80.2 (9.6 s.d.) while the mean SUS score in VR was 75.4 (13.1 s.d.). There was no significant difference between survey scores for the monitor display condition ($p = 0.2535$). The standardized GUESS-18 mean score was 75.3 (11.1 s.d.) and the SUS mean score was 76.9 (11.9 s.d.).

We compared GUESS-18 questions that measure play engrossment to determine if user immersion was affected by display condition. Questions 5 and 6 in the GUESS-18 measure the 'play engrossment' factor, so we performed t-tests on these questions

between display conditions. The VR condition was found to be significantly more engrossing than the monitor ($p < 0.0001$ for both questions). The mean for Q5 in VR was 6.0 (1.1 s.d.) while the mean for Q5 with the monitor was 3.4 (1.8 s.d.), a 55.3% difference. The mean for Q6 in VR was 5.4 (1.9 s.d.) while the mean for Q6 with the monitor was 3.8 (1.9 s.d.), a 34.8% difference.

All questions of the GUESS-18 were compared to determine differences based on display condition.

Table 3. Comparisons for each GUESS-18 question between display condition

GUESS-18 Question	VR Mean	Monitor Mean	P-value
Q1	6.0 (0.7 s.d.)	6.3 (0.8 s.d.)	0.0842
Q2	5.8 (1.2 s.d.)	6.2 (0.8 s.d.)	0.0449*
Q3	5.3 (1.3 s.d.)	5.2 (1.5 s.d.)	0.3827
Q4	5.4 (1.4 s.d.)	5.1 (1.5 s.d.)	0.1756
Q5	6.0 (1.1 s.d.)	3.4 (1.8 s.d.)	< 0.0001*
Q6	5.4 (1.9 s.d.)	3.8 (1.9 s.d.)	< 0.0001*
Q7	6.3 (1.1 s.d.)	6.0 (1.0 s.d.)	0.0758
Q8	2.1 (1.3 s.d.)	2.4 (1.3 s.d.)	0.2150
Q9	5.1 (1.6 s.d.)	4.7 (1.4 s.d.)	0.0894
Q10	4.7 (1.4 s.d.)	4.7 (1.5 s.d.)	0.5
Q11	6.1 (1.1 s.d.)	6.0 (1.1 s.d.)	0.2234
Q12	6.5 (0.6 s.d.)	6.1 (1.0 s.d.)	0.0057*
Q13	6.7 (0.5 s.d.)	6.2 (0.8 s.d.)	0.0001*
Q14	6.7 (0.6 s.d.)	6.5 (0.7 s.d.)	0.0368*
Q15	3.5 (1.7 s.d.)	3.8 (1.5 s.d.)	0.1849
Q16	5.5 (1.6 s.d.)	5.2 (1.6 s.d.)	0.1484
Q17	5.5 (1.5 s.d.)	5.7 (1.3 s.d.)	0.2516
Q18	5.5 (1.5 s.d.)	5.6 (1.4 s.d.)	0.4151

We see that Questions 2, 12,13, and 14 are also significant. Question 2 asks about the game's interface, which could be a product of the game's Head-Up Display (HUD), which was the only display the user saw during testing. Question 12 asks if the game's audio enhances the user's gaming experience, which is greater in VR. Questions 13 and 14 target the personal gratification factor, with both questions indicating greater satisfaction in the VR condition. Questions 13 asks the user if they want to perform as well as possible in the game. Question 14 asks if the user is focused on their own performance while playing the game.

Confirmatory Factor Analysis for the GUESS-18 in VR

We performed CFA with 9 factors to validate the GUESS-18 in its measurements of VR usability. We note that we combined GUESS-18 responses between conditions for a total of 90 survey responses in the CFA.

The rotated factor loading table below represents the loadings between each factor and each question from the GUESS-18 based on the 90 samples. The rotation was performed with the orthogonal varimax technique in JMP. Values of 0.3, 0.6, and 0.9 represent weak, medium, and strong factor loadings, respectively (Briggs & MacCallum, 2003).

Table 4. Factor loadings for the GUESS-19 questions onto the 9 factors. Entries in bold indicate medium or strong factor loading between that GUESS-18 question and each of the nine factors (rounded to the nearest thousandth).

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
Q1	-0.054	0.135	0.052	-0.186	0.090	0.329	**0.817**	-0.107	0.047
Q2	0.247	0.093	0.100	0.009	-0.128	-0.137	**0.844**	0.117	-0.187
Q3	0.308	0.268	0.340	0.107	-0.039	0.574	0.235	-0.152	-0.178
Q4	0.210	0.276	0.324	0.173	0.074	**0.749**	0.027	0.111	-0.044
Q5	-0.040	0.126	0.076	**0.878**	0.189	0.057	-0.157	-0.017	0.058
Q6	-0.026	0.129	0.143	**0.889**	0.096	0.077	0.012	-0.082	-0.121
Q7	0.136	0.574	0.238	0.039	0.067	0.429	0.084	0.070	-0.442
Q8	-0.130	-0.254	-0.230	-0.055	-0.093	-0.095	-0.112	-0.111	**0.866**
Q9	0.182	0.124	**0.832**	0.223	0.107	0.176	0.072	0.0440	-0.212
Q10	0.092	0.231	**0.892**	0.044	0.049	0.173	0.071	0.030	-0.081
Q11	0.155	**0.822**	0.193	0.048	-0.055	0.257	0.070	0.102	-0.235
Q12	0.115	**0.872**	0.142	0.252	0.003	0.027	0.139	0.071	-0.030
Q13	-0.012	0.079	0.045	0.326	**0.834**	0.006	-0.068	0.069	0.041
Q14	0.032	-0.087	0.078	0.010	**0.922**	0.007	0.025	0.016	-0.122
Q15	0.454	-0.057	-0.033	-0.075	-0.241	0.479	0.055	0.539	-0.104
Q16	0.037	0.155	0.060	-0.065	0.126	-0.013	-0.005	**0.914**	-0.074
Q17	**0.929**	0.119	0.103	-0.033	0.012	0.158	0.093	0.039	-0.126
Q18	**0.920**	0.158	0.161	-0.026	0.040	0.120	0.067	0.071	-0.022

Observe that the majority of medium and strong factor loadings come in pairs (e.g., Q13 and Q14 are strongly related to Factor 5). This is expected since the GUESS-18 was designed such that each of the nine factors was measured with two questions. The outliers to this observation are Q3, Q7, and Q15, which have no medium or strong factor loadings with any factor. From this table, each factor is identifiable from our hypothesized model based on the pairs of questions.

Table 5. Factor loadings for GUESS-18 questions onto their corresponding subscales

Factor 1	Q17, Q18	Visual Aesthetics
Factor 2	Q11, Q12	Audio Aesthetics
Factor 3	Q9, Q10	Creative Freedom
Factor 4	Q5, Q6	Play Engrossment
Factor 5	Q13, Q14	Personal Gratification
Factor 6	Q4	Narratives
Factor 7	Q1, Q2	Usability/Playability
Factor 8	Q16	Social Connectivity
Factor 9	Q8	Enjoyment

Two chi-square goodness of fit tests were performed to determine if the number of factors captured by the model was sufficient. At least two of the factors in the model are related, suggesting that the GUESS-18 has overlapping latent factors ($p < 0.0001$). The other chi-square goodness of fit test determined that there are no missing factors in the model, removing the need to identify additional latent factors ($p = 0.7902$).

CFI and RMSEA values were computed in R to determine appropriate model fit. There were 99 degrees of freedom in the observed model. A CFI value of 0.967 was calculated and a RMSEA value of 0.056 was calculated. The CFI evaluates how well the observed model compares to a baseline model in which none of the variables are correlated. Our CFI value indicates that the observed model greatly differs from a null model, suggesting adequate model fit. The RMSEA measures the difference between the observed model and the baseline (null) model per degrees of freedom. Our RMSEA value indicates a discrepancy between the observed model and the baseline model, suggesting adequate model fit. We determined internal consistency with Cronbach's alpha, which determines how closely related a set of items are as a group. Our data yielded a Cronbach's alpha value of 0.776, pointing towards a sufficiently reliable scale.

Table 6. Summary of validation metrics

Model	**Degrees of Freedom**	**Chi-Square**	**CFI**	**RMSEA**	**Cronbach's Alpha**
GUESS-18	99	126.653	0.967	0.056	0.776

Based on the CFA, the GUESS-18 seems to be an effective tool for measuring usability of VR video games. The CFI and RMSEA values indicate excellent model fit, and our Cronbach's alpha value shows sufficient reliability. The factor loadings table reveals at least 1 medium or strong loading for most GUESS-18 questions, although some questions with weaker loadings could be removed without sacrificing the integrity of the model. Playing the game in VR resulted in higher GUESS-18 scores, which translates to greater usability. One performance measure, best lap times, was lower in VR than the monitor, while the remaining performance measures found no difference. Furthermore,

the VR condition stimulated greater engagement, personal gratification, and audio aesthetics than the monitor. On the other hand, the monitor was perceived to have a better interface than VR. Greater levels of motion sickness were also present after interacting with the VR condition. The GUESS-18 provides a better measurement of VR video game usability than the SUS because of its significant differences between display conditions and its validation.

Discussion

The objective of this study was to determine if the GUESS-18 would sufficiently measure the usability of a video game in VR. It was expected that the tool would be sufficiently valid in its measurements and that the VR display condition would lead to superior usability scores compared to the monitor condition. It was also expected that performance would be unchanged between display conditions and that VR would provide greater engagement, immersion, and presence than the monitor condition. This was tested using state-of-the art VR technology in a controlled environment where participants played a racing video game with both VR and monitor displays. CFA determined that the GUESS-18 provided an acceptable model when measuring the usability of video games in VR.

Usability

We hypothesized that the VR condition would yield higher scores from the usability surveys than the monitor condition. Results from the GUESS-18 indicate that VR was significantly more usable than the monitor condition by 6.4% when playing Assetto Corsa. However, there was no significant difference found in SUS scores between display conditions. This is an interesting result because the SUS is widely viewed as the gold standard in system usability, but it did not find a significant difference when the GUESS-18 did. This could be a result of the specificity provided by the GUESS-18's questions that are tailored to common video game factors. The SUS uses broader items that focus on perceived usability, but that generality misses out on the specific factors measured by the GUESS-18.

We hypothesized that there would be no difference in standardized survey scores for the GUESS-18 and SUS between display conditions. Interestingly, there was a significant difference found in which the mean standardized GUESS-18 score was higher than the mean SUS scores by 6.2%. There was no difference between SUS and GUESS-18 scores for the monitor display condition. For all conditions, survey scores were higher than 68, indicating satisfactory usability scores. Since we determined that the GUESS-18 is sufficiently valid in its measurements of VR video game usability and also found a significant difference between display conditions with the GUESS-18, we claim that the GUESS-18 provides a better measurement of VR video game usability than the SUS. At the present, no research has been conducted for comparing scores between the GUESS, GUESS-18, and other usability tool.

It was hypothesized that performance measures between conditions would be unchanged. No differences were found for minor/major errors between display conditions. However, the best lap times from the VR condition were 5% faster than times from the monitor condition. This performance difference could stem from the heightened sense of immersion from the VR condition. Other research has found that greater immersion in VR may lead to better physical performance for activities such as exercise (Kim & Biocca, 2018).

We hypothesized that the VR condition would stimulate greater user engagement, immersion, and presence than the monitor condition (Schutte & Stilinović, 2017, Pallavicini et al., 2019). This was measured with questions 5 and 6 from the GUESS-18, which target the ‘play engrossment’ factor of video games. Results showed that the VR condition was significantly higher than the monitor condition for both questions. This could be a result of the sense insertion into the digital world provided by VR. Not only are users inserted into this world, but the physical world is removed from the user’s sight. By omitting the physical world from the user’s experience comes greater levels of immersion when interacting with the digital medium. Our study found that VR led to greater user engagement than the monitor. This could be caused by blocking outside stimuli with an HMD, leading to more immersive experiences (Genie, 2020).

Significant differences were also found between display conditions for questions 2, 12, 13, and 14 of the GUESS-18. Question 2 asks the user if they found the game’s interface easy to navigate. This result is particularly interesting because participants did not interact with the game’s menus during the experiment. However, the game’s HUD was present in both conditions, suggesting that the HUD was better received in the monitor condition. Question 12 asks if the game’s audio enhances the user’s gaming experience, which was greater in VR. The HMD used in the experiment placed speakers immediately on the user’s ears, while the monitor condition used an external speaker than rested farther away from the user. We believe that the observed difference was caused by the distance between the user’s ears and the source of sound. If the monitor condition implemented headphones, it is expected that no difference would be present. Finally, questions 13 and 14 target the personal gratification factor of the GUESS-18, in which VR scores were notably higher. This result could stem from the user’s inexperience with VR. Since the technology is still on the rise, most users had little to no experience with the technology. Playing in a new and unique environment could lead to a more gratifying user experience while playing the game.

Simulator Sickness

Playing Assetto Corsa in VR led to a significant increase in motion sickness by 282.5%. This drastic increase in motion sickness highlights the shortcomings of modern VR technology and further identifies the need for standardized VR usability tools. SSQ scores higher than 20 indicate a need for improvement (Kennedy et al., 1993). Since the average SSQ score after engaging with VR was 24.1, there is a clear need for usability improvements within the observed system. The increase in SSQ scores after playing in VR can be explained by sensory conflict theory, which explains that a conflict with the body’s senses can lead to motion sickness. This experiment had users drive a car in VR, so there was a conflict with users’ sight and touch, leading to motion sickness. These SSQ scores seem to be quite high when compared to other studies involving VR. A study implementing Sony’s VR system across several games found that 58% of participants had low SSQ scores (0 to 8), 33% had medium scores (9 to 20), and 9% had high scores (21 and above) (Norman, 2018). Bruck and Watters found that 10 of the 16 items of the SSQ were significantly higher in high simulated motion tasks compared to low simulated tasks (2009). Another study used an HMD for augmented

reality and found average SSQ scores of 8.65 across various conditions and environments (Vovk et al., 2018).

VR research has repeatedly shown moderate to high SSQ scores after using a VR system (Dużmańska et al., 2018). Several theories have been introduced to explain this phenomenon. The first and foremost explanation is sensory conflict theory (Reason & Brand, 1975). Sensory conflict theory explains that the dissonance between difference senses of the body can lead to motion sickness. When a user engages with VR, there is a disconnection in what the user sees on the screen and what they are feeling on their body. This disconnection causes a conflict in the brain because of the mismatched information, leading to simulator sickness.

Another common theory when explain simulator sickness is the Postural Instability Theory (Riccio & Stoffregen, 1991). This theory suggests that motion sickness occurs when an individual is exposed to long-lasting postural instability without learning how to adapt to that environment with proper posture or balance. The classic example for this theory is traveling on a ship, where one may struggle with balance before adapting to the sway of the ship, leading to seasickness. Our study had users drive a car at high speeds in VR, where they saw movement on the display but did not physically feel the movement as they drove. This sensory conflict could be the cause of higher SSQ scores in the VR condition. There was no postural sway in the experiment, ruling out that possibility for the increased motion sickness.

Validation

We hypothesized that the GUESS-18 would be valid in its measurements of VR usability and the results support this hypothesis. The CFI value of 0.967 is above the acceptable threshold of 0.9, pointing to satisfactory model fit. Additionally, the RMSEA value of 0.056 is also in the acceptable range since it is below 0.06. The Questionnaire is internally consistent as determined by a Cronbach's alpha value of 0.776, above the minimum of 0.7. We obtained a chi-square value of 126.653. Higher chi-square values usually indicate poor model fit, but chi-square values are inflated as sample sizes increase. Thus, CFI and RMSEA values are better indicators for model fit when validating a questionnaire. The developers of the GUESS-18 found similar results with their initial validation of the tool (chi-square = 137.0, CFI = 0.974, RMSEA = 0.043 (Keebler et al., 2020)). Other than the initial validation study, no research has been conducted in which the GUESS or GUESS-18 was validated with EFA or CFA. However, a recent study used the GUESS to develop a framework for measuring usability of an educational game (Atmaja & Sugiarto, 2021). This study found sufficient reliability of the framework with Cronbach's alpha values ranging from 0.661 to 0.868 depending on the subscale from the GUESS-18. They did not perform EFA or CFA in their validation.

A chi-square goodness of fit test was performed to determine if the hypothesized model fit the observed data appropriately. It was determined that at least two of the factors in the model are related, suggesting further reduction of the GUESS-18 could lead to a more optimized usability tool for measuring VR video games. Questions 3, 7, and 15 did

not provide a medium or strong factor loading across the 9 factors, suggesting that their removal could produce a more optimized tool for measuring usability of video games in VR. Furthermore, it was determined that there were no missing latent factors in the model.

The GUESS-18 provides valid measurements for usability of the VR video game Assetto Corsa. After all, VR video games are still video games, and since the GUESS-18 was determined to be an effective tool for measuring usability of video games (Keebler et al., 2020), these results were anticipated. While the GUESS-18 was found to be an effective tool for measuring usability of Assetto Corsa in VR, future research could perform CFA for other VR games to generalize the results obtained from the present study.

Relevance to Modern Research

Research has been conducted to understand how a player's experience is altered when playing a game in VR. Pallavicini et al. performed an experiment with immersive and non-immersive modalities to determine usability, emotional response, and presence from users (2019). They found that no differences between immersive and non-immersive conditions for usability and performance scores. However, the perceived sense of presence was higher in VR than a monitor. Additionally, players exhibited greater emotional responses after playing in VR. Our study found that VR was more engaging and immersive than the monitor display condition, so we could compare our results similarly as immersive (VR) and non-immersive (monitor). Our immersive condition found differences in usability and one performance measure (speed), contradicting the previous work. Emotional responses were not measured in either condition.

Previous research has found conflicting results with VR gaming and levels of player satisfaction. Shelstad et. al found greater player satisfaction when playing a tower-defense strategy game in VR than on a monitor (2017). Results indicated higher levels of engrossment, creative freedom, visual/auditory aesthetics, and general enjoyment when playing in VR. VR displays have also been found to be more engaging and provide a greater sense of presence than monitors for aircraft safety procedures (Buttussi & Chittaro, 2018). There is evidence suggesting VR elicits empathy and engagement more so than traditional 2D displays (Schutte & Stilinović, 2017). Our study found that personal gratification, immersion, and audio aesthetics was significantly higher in VR, aligning with these past results. However, other research has found little difference between VR and monitor displays for video game satisfaction. One study found no difference with video game satisfaction between VR and monitor displays (Yildirim et al., 2018). Another study found similar results; video game satisfaction was unchanged between display mediums (Carroll et al., 2019). Both of these studies contradict the findings of the present study.

In April of 2022, an article was published that addressed the effects of mixed reality on video game satisfaction using the GUESS-18 and ENJOY for the popular mobile game Angry Birds (Sinlapanuntakul et al., 2022). They implemented a strategy game in mixed

reality and on a mobile device that used survey responses to measure satisfaction, enjoyment, and performance when using a mixed reality medium. They found a significant difference in each of these measures, with satisfaction and enjoyment being higher in mixed reality than on a mobile device. However, user performance was higher in the mobile condition than the mixed reality condition. This study found no significant difference in SSQ scores after using the mixed reality configuration. While mixed reality and VR differ in their own ways, they both implement technology that enhances displays for humans. This mixed reality study observes similar findings to our work, where satisfaction and personal gratification are higher in their respective alternative reality conditions. The performance decrease in the mixed reality study is a product of the users' level of comfort with the gaming medium. Since most participants of the study were more familiar with the game on mobile, it is understandable that their performance was better on mobile, where they had more experience with the game. The difference in SSQ scores could be an indicator that current technology in mixed reality leads to less motion sickness than modern VR systems.

With the development of any technology comes the need for standardized tools that measure the quality of its development. VR technology has come a long way over the years, but still lacks important features to provide an optimal user experience. There is currently no validated questionnaire centered around perceived usability of a VR. However, the GUESS-18 provides a satisfactory method for measuring usability of video games in VR. The CFA showed acceptable levels of model fit and reliability. Additionally, participants preferred playing in VR more than a traditional monitor. These findings emphasize the potential VR has to offer, albeit with some usability obstacles to overcome, like motion sickness.

Limitations and Future Research

As with any research, there are limitations that must be considered to fully quantify the quality of work. For validating questionnaires, 10 samples per factor is typically recommended (Samuels, 2016). This study obtained 90 samples, which provides the bare minimum as suggested. Future studies that use the GUESS-18 for VR usability should incorporate a larger sample size.

This validation was performed using only one VR video game. Future research could continue testing the GUESS-18 across other VR video games to broaden the scope of the tool's validity. This could further validate the GUESS-18 for general VR video game usability.

Certain factors measured by the GUESS-18 were not prominent in the experiment. (e.g., the narratives factor asks about a game's story, but there was no story component present in the experiment). Future research could test a VR game that thoroughly incorporates each factor measured from the GUESS-18.

Another concern is the length of the testing sessions. Participants were only tested for 3 minutes in each condition for a total of 6 minutes of testing across two display types. Usability studies are concerned with learnability, performance, and errors. Short testing

sessions could prevent users from becoming familiar with the game, leading to suboptimal performance.

Both display conditions were tested during the same session. This could influence user performance, as it was determined that users had significantly higher levels of motion sickness after the VR condition, which could carry over into performance of the monitor condition for those who started in VR.

Our experiment had participants complete the SSQ before and after playing VR, but every other participant used the monitor before VR. This could have influenced the SSQ survey data since using the monitor can lead to increased motion sickness.

Due to these limitations, certain aspects of the GUESS-18 may need to be further testing to determine general validation. We determined that the narratives and social connectivity subscales of the GUESS-18 were not present in the experiment, suggesting those components should be present in another validation study. Furthermore, a longer testing session could provide a better usability experience and more accurate data. Finally, a future experiment could give participants a specific objective or goal to motivate users further.

Despite the validity of the GUESS-18 for measuring usability of VR video games, there is a greater need in the world of general VR usability; There is currently no standardized tool for measuring general VR usability. Further research is needed for the development and validation of a general VR usability tool that measures physical usability factors present with VR HMDs such as motion sickness, HMD pressure, and controllers. The tool should also be capable of measuring digital factors such as VR interface, engrossment, visual/audio aesthetics, and general usability. The development of such a tool is imperative for the usability of VR technology as it develops in the near future.

Conclusion

This study performed confirmatory factor analysis on the GUESS-18 for measuring the usability of a virtual reality racing game. Results found that the GUESS-18 is an effective measurement tool when determining the usability of the video game in VR. These results were obtained through an acceptable CFI value of 0.967 and RMSEA value of 0.056. Our Cronbach's alpha value of 0.776 indicates sufficient reliability among questionnaire items. This study emphasizes strong construct validity and content reliability for the GUESS-18. CFA is a common method for testing construct validity, which determines if the model is measuring what it was designed to measure. Our findings indicate that the GUESS-18 provides strong construct validity. However, other types of validity must be considered when determining the effectiveness of a questionnaire. Content validity refers to a test's representation of all factors of the construct. With the GUESS-18, this refers to the 9 subscales measured by 18 questions. Our study found strong content validity through the chi-square goodness of fit tests, which found that there were no latent factors missing in the observed model. Internal and external validity should also be considered. Internal validity asks if the relationships tested are reliable and not influenced by outside factors. External validity focuses on the applicability to generalization to other events or situations. Typically, increasing one of these leads to a decrease in the other, so we must consider which of these validity types was prioritized in the study. Our study has higher internal validity because of the relationships found in the laboratory setting. To evaluate external validity, the GUESS-18 could be applied to a variety of VR games from people of all ages to better understand how well the tool generalizes. Additionally, GUESS-18 scores indicated that participants preferred the game when playing in VR as opposed to a traditional monitor. One performance measure reflected this with lap times, which were faster in the VR condition. All other performance measures showed no difference between display conditions.

This was the first study to validate a questionnaire that measures usability of VR video games, filling a gap in the larger hole of VR usability. It was also the first study to compare scores between the GUESS-18 and SUS usability tools and determined that the GUESS-18 provides a better method for measuring VR video game usability than the SUS. These findings have many beneficial applications beyond the scope of research. Future VR video game developers can use the GUESS-18 to obtain usability scores for their games during development, with each subscale identifying the shortcomings of the product. Furthermore, the GUESS-18 has potential to accurately measure VR software. As the technology advances over the coming years, developers must understand how usable their products are to construct effective systems. VR has been shown to be applicable in a variety of different ways, but it must make strides in usability to fully realize its potential.

Appendix A. GUESS-18 Questionnaire

This survey is designed to measure the usability/playability of video games. It is measured on a 7-point Likert scale (strongly disagree, disagree, somewhat disagree, neutral, somewhat agree, agree, and strongly agree).

Usability/Playability

1. I find the controls of the game to be straightforward.
2. I find the game's interface to be easy to navigate.

Narratives

3. I am captivated by the game's story from the beginning.
4. I enjoy the fantasy or story provided by the game.

Play Engrossment

5. I feel detached from the outside world while playing the game.
6. I do not care to check events that are happening in the real world during the game.

Enjoyment

7. I think the game is fun.
8. I feel bored while playing the game.

Creative Freedom

9. I feel the game allows me to be imaginative.
10. I feel creative while playing the game.

Audio Aesthetics

11. I enjoy the sound effects in the game.
12. I feel the game's audio enhances my gaming experience.

Personal Gratification

13. I want to do as well as possible during the game.
14. I am very focused on my own performance while playing the game.

Social Connectivity

15. I find the game supports social interaction (e.g., chat) between players.
16. I like to play this game with other players.

Visual Aesthetics

17. I enjoy the game's graphics.
18. I think the game is visually appealing.

Appendix B. SUS Questionnaire

This survey is designed to measure general usability of a system. It is measured on a 5-point Likert scale (Strongly disagree, disagree, neutral, agree, and strongly agree).

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with the system.

Appendix C. Simulator Sickness Questionnaire

This survey is meant to capture motion sickness of an individual. Participants will complete this questionnaire immediately before and after exposure to VR. Participants respond to the following symptoms with one of four responses (none, slight, moderate, severe). SSQ Scores are obtained used using the scoring table given below.

1. General Discomfort
2. Fatigue
3. Headache
4. Eye strain
5. Difficulty focusing
6. Increased salivation
7. Sweating
8. Nausea
9. Difficulty Concentrating
10. “Fullness of the Head”
11. Blurred Vision
12. Dizzy (eyes open)
13. Dizzy (eyes closed)
14. Vertigo
15. Stomach awareness
16. Burping

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

$$N = [1] \times 9.54$$
$$O = [2] \times 7.58$$
$$D = [3] \times 13.92$$
$$TS = ([1] + [2] + [3]) \times 3.74$$

Appendix D. IRB Forms and Survey Links

IRB Consent Form

Consent Form for Non-Clinical Study

Title: Evaluating the Usability of a Virtual Reality Racing Game using the GUESS-18

Purpose of the study: The purpose of this study is to administer and validate the Game User Experience Satisfaction Scale-18 (GUESS-18) for measuring virtual reality usability with Head-Mounted Displays.

Study Procedures: You will engage in the racing video game Assetto Corsa. You will use the Hotlap feature of the game, where you will complete laps of a racetrack as fast as possible with no other vehicles present. We will save the recordings for your testing session using the game's features of saving replays. You will use the HTC VIVE Pro Eye head-mounted display from Patrick F. Tayler Hall room 2348. You will use the Logitech G920 Driving Force Steering Wheel and Floor Pedals to control the game. For the 2D display, a traditional monitor is used.

You will be tested with two display types: a VR headset and 2D monitor. You will be given a practice session before testing in which you should become familiar with the controls. Practice sessions will be held for each display type, totaling 2 practice sessions in this session. After finishing, you will complete a survey with Qualtrics, a web-based survey tool. The survey consists of the GUESS-18, System Usability Scale, and Simulator Sickness Questionnaire. You will complete one survey for each display type, totaling 2 surveys per participant. Testing concludes after both surveys are submitted.

Risks/Discomforts: Exposure to virtual reality can cause visually induced motion sickness (VIMS). Common symptoms of VIMS include nausea, vomiting, dizziness, disorientation, headaches, sweating, blurred vision, fatigue, or loss of balance. These symptoms are short-term and reversible. You will be seated during testing to reduce the risk of injury.

COVID-19 Mitigation: Please reschedule any testing plans if you are experiencing sickness symptoms. Face coverings are required at all times during testing. Testing stations are sanitized between each participant. Social distancing of 6ft will be observed at all times.

Benefits: There are no direct benefits; however, this experiment may provide future information that is helpful for improving understanding of usability testing with virtual reality environments. Other than the extra credit offered as compensation, participation or withdrawal from the experiment will have no impact on regular grading activates during the course. At the completion of the sessions, you will receive 0.5 points added to your final grade for your Industrial Engineering Course with your professor's approval. Should you choose to withdraw, extra credit will not be awarded for the experiment.

However, you will have the option to earn extra credit in your Industrial Engineering Course through other activities, which are posted on the course Moodle page.

Right to Refuse: You have the right to withdraw from participation at any time. There are no penalties for withdrawal except that extra credit will not be awarded in your class. If you refuse to comply with the objectives of the study, you will be removed from the premises by an investigator.

Performance Sites: 2348 Patrick F. Taylor Hall, Baton Rouge, LA 70803

Number of Participants: 40 participants will be included in this study.

Investigators: The following are the investigators for the study:

Laura Ikuma, PhD (likuma@lsu.edu), 225-588-9715

Drew Carman (wcarma1@lsu.edu), 205-706-1782

Subject Inclusion: LSU students will be recruited for this study. You must have normal or corrected to normal vision and full motor control of your body to participate. You must also have no cognitive disabilities that would impair your ability to drive a car. If you are under the age of 18, you are not permitted to join the study.

Exclusion Criteria: Any individual with low or uncorrected vision or lacking full motor control of their body is not accepted for testing. Those of age 17 and below are excluded from participation. If you possess a cognitive disability that impairs your performance of driving a car, you are not permitted to join the study.

Privacy: The LSU Institutional Review Board (oversees university research with human participants) may inspect and/or copy records of the study. Results may be published, but no names or identifying information will be released in the publication. Other than these circumstances, participant identity will remain confidential unless legal disclosure is required.

Principal Investigators:

Laura Ikuma, PhD (likuma@lsu.edu), 225-588-9715

Signatures: This research has been explained to me and my questions have been answered about participation. I understand the risks of participation and that I may withdraw at any time for any reason. If I have questions about my rights or other issues, I can contact Alex Cohen, Institutional Review Board, (225) 578-8692, irb@lsu.edu, or www.lsu.edu/research. I agree to participate in the study as described above and understand the investigator’s obligation to provide me with a copy of the signed consent form.

______________________________ ______________________________

Participant Signature Date

____

Printed Name

IRB Approval

LSU | Office of Research & Economic Development

TO:	Laura H Ikuma LSUAM | Col of ENGR | MECH and IE - Industrial Engineering | CC00178
FROM:	Alex Cohen Chairman, Institutional Review Board
DATE:	29-Mar-2022
RE:	IRBAM-22-0217
TITLE:	Evaluating the Usability of a Virtual Reality Racing Game using the GUESS-18
SUBMISSION TYPE:	Initial Application
Review Type:	Expedited Review
Risk Factor:	Minimal
Review Date:	28-Mar-2022
Status:	Approved
Approval Date:	28-Mar-2022
Approval Expiration Date:	27-Mar-2023
Expedited Categories:	07
Requesting Waiver of Informed Consent:	No
Re-review frequency:	Annually
Number of subjects approved:	40
LSU Proposal Number:	
By:	Alex Cohen, Chairman

Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.

5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
7. Notification of the IRB of a serious compliance failure.
8. **SPECIAL NOTE: When emailing more than one recipient, make sure you use bcc. Approvals will automatically be closed by the IRB on the expiration date unless the PI requests a continuation.**

** All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/research*

Louisiana State University
131 David Boyd Hall
Baton Rouge, LA 70803

O 225-578-5833
F 225-578-5983
http://www.lsu.edu/research

IRB Amended Approval

LSU | Office of Research & Economic Development

TO:	Laura H Ikuma LSUAM | Col of ENGR | MECH and IE - Industrial Engineering | CC00178
FROM:	Alex Cohen Chairman, Institutional Review Board
DATE:	18-Apr-2022
RE:	IRBAM-22-0217
TITLE:	Evaluating the Usability of a Virtual Reality Racing Game using the GUESS-18
New Protocol/Amendment/Continuation:	Amendment
Brief Amendment Description:	We are requesting to increase the number of participants from 40 to 60. No other changes are requested. We have changed the number in the consent form and in the application.
Review Type:	Expedited Review
Risk Factor:	Minimal
Review Date:	16-Apr-2022
Status:	Approved
Approval Date:	16-Apr-2022
Approval Expiration Date:	27-Mar-2023
Re-review frequency:	(annual unless otherwise stated)
Number of subjects approved:	60
By:	Alex Cohen, Chairman

Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the

study ends.

5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
7. Notification of the IRB of a serious compliance failure.
8. **SPECIAL NOTE: When emailing more than one recipient, make sure you use bcc.**

** All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/research*

Louisiana State University
131 David Boyd Hall
Baton Rouge, LA 70803

O 225-578-5833
F 225-578-5983
http://www.lsu.edu/research

Survey Links

The links below direct to the instruments we used in Qualtrics.

Consent Form

Demographics Survey

Simulator Sickness Questionnaire

Usability Testing Survey

References

Arh, T., & Blažič, B. J. (2008). A Case Study of Usability Testing - the SUMI Evaluation Approach of the EducaNext Portal. *WSEAS Transactions on Information Science & Applications*, *5*(2), 175–181.

Atmaja, P. W., & Sugiarto, S. (2021). Balancing Entertainment, Cost, and Educational Strength: A Design Framework for Medium-Coupling Educational Games. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, *4*, 27–40. https://doi.org/10.22219/kinetik.v6i1.1158

Baker, H. (2021). *Beat Saber Passes $100 Million In Revenue On Quest Platform Alone*. https://uploadvr.com/beat-saber-100-million-revenue/#:~:text=In February%2C Facebook revealed that,Store (Quest and Rift).

Bateye, O., Fatoye, F. A., Adejumobi, A. S., Olatoye, F. S., Makinde, M. O., Mbada, C. E., Salami, A. J., Dada, O. O., Gambo, I. P., & Odole, A. C. (2020). Development and feasibility testing of a back extension-glide virtual reality game for low-back pain. *International Journal of Telemedicine and Clinical Practices*, *3*(3), 243. https://doi.org/10.1504/ijtmcp.2020.10026527

Bente, G., & Breuer, J. (2010). Why so serious? On the Relation of Serious Games and Learning. *Eludamos - Journal for Computer Game Culture*, *4*(1), 7–24. http://www.eludamos.org/index.php/eludamos/article/viewArticle/vol4no1-2

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, *38*(1), 25–56. https://doi.org/10.1207/S15327906MBR3801_2

Brooke, J. (1986). SUS: A “Quick and Dirty” Usability Scale. *Usability Evaluation In Industry*, 207–212. https://doi.org/10.1201/9781498710411-35

Bruck, S., & Watters, P. (2009). *Estimating Cybersickness of Simulated Motion Using the Simulator Sickness Questionnaire (SSQ): A Controlled Study*. https://doi.org/10.1109/CGIV.2009.83

Buttussi, F., & Chittaro, L. (2018). *Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario*. *24*(2), 1063–1076.

Carroll, M., Osborne, E., & Yildirim, C. (2019). Effects of VR gaming and game genre on player experience. *2019 IEEE Games, Entertainment, Media Conference, GEM 2019*. https://doi.org/10.1109/GEM.2019.8811554

Child, D. (1990). The Essential of Factor Analysis: Second Edition. *London: Cassel Educational Limited*.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Conference on Human Factors in Computing Systems - Proceedings*, *Part F1302*, 213–218. https://doi.org/10.1145/57167.57203

Dużmańska, N., Strojny, P., & Strojny, A. (2018). *Can Simulator Sickness Be Avoided? A Review on Temporal Aspects of Simulator Sickness*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02132/full

Fang, Y., & Lin, C. (2019). The Usability Testing of VR Interface for Tourism Apps. *Applied Sciences*, *9*(16), 3215. https://doi.org/10.3390/app9163215

Fox, J., Pittaway, L., & Uzuegbunam, I. (2018). Simulations in Entrepreneurship Education: Serious Games and Learning Through Play. *Entrepreneurship Education and Pedagogy*, *1*(1), 61–89. https://doi.org/10.1177/2515127417737285

Fussell, S. G., Derby, J. L., Smith, J. K., Shelstad, W. J., Benedict, J. D., Chaparro, B. S., Thomas, R., & Dattel, A. R. (2019). Usability Testing of a Virtual Reality Tutorial. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2303–2307. https://doi.org/10.1177/1071181319631494

Fussell, S. G., & Hight, M. P. (2021). Usability Testing of a VR Flight Training Program. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 1124–1128. https://doi.org/10.1177/1071181321651096

Genie, G. M. (2020). *What Does Mixed Reality Mean For the Video Game Industry?* https://www.gamemarketinggenie.com/blog/mixed-reality-video-game-industry

Haleem, A., & Javaid, M. (2018). Role of CT and MRI in the design and development of orthopaedic model using additive manufacturing. *Journal of Clinical Orthopaedics and Trauma*, *9*(3), 213–217. https://doi.org/10.1016/j.jcot.2018.07.002

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Javaid, M., & Haleem, A. (2020). Virtual reality applications toward medical field. *Clinical Epidemiology and Global Health*, *8*(2), 600–605. https://doi.org/10.1016/j.cegh.2019.12.010

Keebler, J., Shelstad, W., Smith, D., Chaparro, B., & Phan, M. (2020). Validation of the GUESS-18: A Short Version of the Game User Experience Satisfaction Scale (GUESS). *Journal of Usability Studies*, *16*(1), 49.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, *3*(3).

Khooshabeh, P., Choromanski, I., Neubauer, C., Krum, D. M., Spicer, R., & Campbell, J. (2017). Mixed reality training for tank platoon leader communication skills. *2017 IEEE Virtual Reality (VR)*, 333–334. https://doi.org/10.1109/VR.2017.7892312

Kim, G., & Biocca, F. (2018). *Immersion in Virtual Reality Can Increase Exercise Motivation and Physical Performance* (pp. 94–102). https://doi.org/10.1007/978-3-319-91584-5_8

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction*, *14*(3–4), 463–488. https://doi.org/10.1207/s15327590ijhc143&4_11

Lewis, J. R. (2018). *The System Usability Scale : Past , Present , and Future*. *34*(7), 577–590.

Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5619 LNCS*, 94–103. https://doi.org/10.1007/978-3-642-02806-9_12

Liang, J., Xian, D., Liu, X., Fu, J., Zhang, X., Tang, B., & Lei, J. (2018). Usability study of mainstream wearable fitness devices: Feature analysis and system usability scale evaluation. *JMIR MHealth and UHealth*, *6*(11). https://doi.org/10.2196/11066

Mahmoud, M. M., & Kamel, M. M. (2010). Using exploratory factor analysis model (EFA) for determination the main factors of train's accidents in Egypt (APPLIED STUDY). *Applied Mathematical Sciences*, *4*(36–39), 1883–1897.

Manero, A., Sparkman, J., Dombrowski, M., Buyssens, R., & Smith, P. A. (2018). Developing and training multi-gestural prosthetic arms. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10909 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-319-91581-4_32

Mann, S., Furness, T., Yuan, Y., Iorio, J., & Wang, Z. (2018). *All Reality: Virtual, Augmented, Mixed (X), Mediated (X,Y), and Multimediated Reality*. *X*. http://arxiv.org/abs/1804.08386

Martin, A., Cross, S., & Attoe, C. (2020). The use of in situ simulation in healthcare education: Current perspectives. *Advances in Medical Education and Practice*, *11*, 893–903. https://doi.org/10.2147/AMEP.S188258

Mittmann, G., Zehetmayer, S., & Schrank, B. (2021). *Study Protocol for a Randomized Controlled Trial to Evaluate the Effectiveness of a Serious Game Targeting Interpersonal Emotion Regulation in Early Adolescents*. 1–24.

Mossel, A., Schoenauer, C., Froeschl, M., Peer, A., & Goellner, J. (2021). Immersive training of first responder squad leaders in untethered virtual reality. *Virtual Reality*, *25*(3), 745–759. https://doi.org/10.1007/s10055-020-00487-x

Mossel, A., Vienna, A.-, Peer, A., & Göllner, J. (2017). Requirements Analysis on a Virtual Reality Training System for CBRN Crisis Preparedness. *Proceedings of the 59th Annual Meeting of the ISSS*, *1*(1), 1–20. file:///C:/Users/ismet/Dropbox/batan2021/CBRN/References/2016 Mossel REQUIREMENTS ANALYSIS ON A VIRTUAL REALITY TRAINING SYSTEM FOR CBRN CRISIS PREPAREDNESS.pdf

Naeini, S., Mostowfi, S., & Naeini, H. S. (2015). Using QUIS as a Measurement Tool for User Satisfaction Evaluation (Case Study: Vending Machine) Fashion design View project Ergonomic Product Design methodology View project Using QUIS as a Measurement Tool for User Satisfaction Evaluation (Case Study: Ve. *International Journal of Information Science*, *2015*(1), 14–23. https://doi.org/10.5923/j.ijis.20150501.03

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Conference*

on Human Factors in Computing Systems - Proceedings, 152–158. https://doi.org/10.1145/191666.191729

Norman, K. L. (2018). Evaluation of virtual reality games: Simulator sickness and human factors. *CEUR Workshop Proceedings*, *2246*.

Pal, D., & Vanijja, V. (2020). Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. *Children and Youth Services Review*, *119*(September), 105535. https://doi.org/10.1016/j.childyouth.2020.105535

Pallavicini, F., & Pepe, A. (2019). Comparing Player Experience in Video Games Played in Virtual Reality or on Desktop Displays. *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, 195–210. https://doi.org/10.1145/3341215.3355736

Pallavicini, F., Pepe, A., & Minissi, E. (2019). *Gaming in Virtual Reality : What Changes in Terms of Usability , Emotional Response and Sense of Presence Compared to Non-Immersive Video Games ?* https://doi.org/10.1177/1046878119831420

Phan, M. H., Keebler, J. R., & Chaparro, B. S. (2016). The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). *Human Factors*, *58*(8), 1217–1247. https://doi.org/10.1177/0018720816669646

Porras, D., Sharon, H., Inzelberg, R., Ziv-Ner, Y., Zeilig, G., & Plotnik, M. (2019). Advanced virtual reality-based rehabilitation of balance and gait in clinical practice. *Therapeutic Advances in Chronic Disease*, *10*, 1–16. https://doi.org/10.1177/2040622319868379

Reason, J. T., & Brand, J. J. (1975). *Motion Sickness*. Oxford: Academic Press.

Riccio, G. E., & Stoffregen, T. A. (1991). An ecological Theory of Motion Sickness and Postural Instability. *Ecological Psychology*, *3*(3), 195–240. https://doi.org/10.1207/s15326969eco0303_2

Samuels, P. (2016). Advice on Reliability Analysis with Small Samples. *Journal ResaerchGate Technical Report*, *August 2015*, 2–6. https://doi.org/10.13140/RG.2.1.1495.5364

Schorer, A., & Protopsaltis, A. (2021). Learners' user experience assessment of a serious game for social innovation education. *Proceedings of the European Conference on Games-Based Learning*, *2021-Septe*(October), 622–630. https://doi.org/10.34190/GBL.21.051

Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and Emotion*, *41*(6), 708–712. https://doi.org/10.1007/s11031-017-9641-7

Sharar, S., Miller, W., Teeley, A., Soltani, M., Hoffman, H., Jensen, M., & Patterson, D. (2008). *Applications of virtual reality for pain management in burn-injured patients*. *8*(11), 1667–1674. https://doi.org/10.1586/14737175.8.11.1667.Applications

Shelstad, W. J., Chaparro, B. S., & Keebler, J. R. (2019). Assessing the User Experience of Video Games: Relationships Between Three Scales. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 1488–1492. https://doi.org/10.1177/1071181319631300

Shelstad, W. J., Hosein, A. A., Keebler, J. R., & Chaparro, B. S. (2020). Using User Experience Scales to Predict Video Game Play and Purchasing Intent. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 2065–2069. https://doi.org/10.1177/1071181320641500

Shelstad, W. J., Smith, D. C., & Chaparro, B. S. (2017). Gaming on the rift: How virtual reality affects game user satisfaction. *Proceedings of the Human Factors and Ergonomics Society*, *2017-Octob*, 2072–2076. https://doi.org/10.1177/1541931213602001

Singh, R., Javaid, M., Kataria, R., Tyagi, M., Haleem, A., & Suman, R. (2020). *Significant applications of virtual reality for COVID-19 pandemic*. *January*.

Sinlapanuntakul, W., Derby, J. L., & Chaparro, B. S. (2022). *Understanding the Effects of Mixed Reality on Video Game Satisfaction , Enjoyment , and Performance*. *53*(3), 237–252. https://doi.org/10.1177/10468781221094473

Sinlapanuntakul, W., Shelstad, W. J., Derby, J. L., & Chaparro, B. S. (2020). *Gameplay in Mixed Reality : How It Differs from Mobile Gameplay*. *61*(1), 2076.

Suhr, D. D. (2006). Exploratory or confirmatory factor analysis? *Rehabilitation*, *45*(4), 243–248. https://doi.org/10.1055/s-2006-940029

Susi, T., Johannesson, M., & Backlund, P. (2007). Serious Games - An Overview. *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2017*, *2018-Janua*, 488–491. https://doi.org/10.1109/UEMCON.2017.8249059

Van Ommen, C., & Chaparro, B. S. (2021). Assessing Video Game Satisfaction of Gamers with Disabilities. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 822–826. https://doi.org/10.1177/1071181321651319

Vovk, A., Wild, F., Guest, W., & Kuula, T. (2018). Simulator sickness in Augmented Reality training using the Microsoft HoloLens. *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*, 1–9. https://doi.org/10.1145/3173574.3173783

Waller, D., Hunt, E., & Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence: Teleoperators and Virtual Environments*, *7*(2), 129–143. https://doi.org/10.1162/105474698565631

Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Labratory Systems*, *40*. https://doi.org/10.1007/978-0-387-87811-9_2

Yildirim, C., Carroll, M., Hufnal, D., Johnson, T., & Pericles, S. (2018). Video Game User Experience: To VR, or Not to VR? *2018 IEEE Games, Entertainment, Media Conference, GEM 2018*, 125–131. https://doi.org/10.1109/GEM.2018.8516542

Yoshimura, B. (2021). *Making Health Education Palatable : A Usability Study of a Digital Game about Nutrition*.

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, *30*(March). https://doi.org/10.1016/j.edurev.2020.100326

Vita

William “Drew” Carman attended the University of Montevallo to obtain his bachelor’s degree in mathematics in December 2018. He then obtained his M.S. in mathematics from LSU in May of 2021. He then moved to the industrial engineering program at LSU, obtaining his M.S. in Industrial Engineering in August of 2022. His hobbies include rock climbing, gaming, and enjoying the great outdoors.