# Space Games:
# Evaluating Game-Based Virtual Reality in Higher Education

*Lana Laskey*[1]

**Abstract**

With increasing global dependence on satellite technology, space traffic has grown exponentially over the last decade. Enhanced education and training of future mission operators will be necessary to meet this growing demand. The complexity of satellite mission operations poses a challenge in education and training. Remote spacecraft are elusive and difficult for a trainee to visualize and involve a steep learning curve. However, the integration of *game-based virtual reality* into spacecraft simulation and training may assist in overcoming these challenges. This research study explored the integration of *game-based virtual reality* into a university course involving spacecraft operations. Virtual spacewalks allowed student participants to conduct visual inspections and interact directly with spacecraft components. The immersive virtual reality environment prolonged cognitive engagement and game mechanics influenced motivation, both cornerstones in learning. After completing the training scenarios, user experience was assessed with several validated scales measuring system usability, user satisfaction, cognitive loading, and any potential simulator sickness. Results revealed satisfactory scores in all categories with minimal simulator sickness. The integrated use of *game-based virtual reality* in the classroom provided an enhanced learning experience in a safe and repeatable environment that might be difficult with traditional teaching methods. This paper will evaluate *game-based virtual reality* when integrated into higher education or other training environments.

**Keywords**

Game-Based Virtual Reality, Instructional Design, Simulation, Spacecraft Operations, Training

[1] Corresponding author: Embry Riddle Aeronautical University, U. S. A., Lana.Laskey@erau.edu

**Acronyms/Abbreviations**

| | |
|---|---|
| *CLS* | *Cognitive Load Scale* |
| *EL* | *Extraneous Cognitive Load* |
| *EL_vr* | *Extraneous Cognitive Load due to Virtual Reality* |
| *GBVR* | *Game-Based Virtual Reality* |
| *GL* | *Germane Cognitive Load* |
| *GUESS* | *Game User Experience and Satisfaction Scale* |
| *IL* | *Intrinsic Cognitive Load* |
| *MCLSVE* | *Multidimensional Cognitive Load Survey for Virtual Environments* |
| *SSQ* | *Simulator Sickness Questionnaire* |
| *SUS* | *System Usability Scale* |
| *VR* | *Virtual Reality* |
| *VRSQ* | *Virtual Reality Sickness Questionnaire* |

## 1. Introduction

Game-based instruction uses game mechanics for serious educational purposes [1] and has been found to increase learner satisfaction and motivation [2]. Additionally, virtual reality (VR) applications provide an immersive learning environment encouraging concentration and prolonged cognitive engagement [3]. These elements are essential for effective learning and have been found helpful in teaching complex disciplines [4]. Merging the two instructional strategies produces an enhanced pedagogical approach known as game-based virtual reality (GBVR). When developing instructional tools and techniques for a complex discipline such as spacecraft operations, employing GBVR may help encourage learner motivation and prolonged cognitive engagement necessary to achieve learning objectives. This study aims to integrate and evaluate an instructional design using GBVR in higher education and provide a quantified pedagogical assessment for educational practitioners, researchers, and industry personnel tasked with training complex disciplines.

## 2. Methodology

A quantitative experimental design was employed to examine the user evaluation of GBVR when integrated into a university course. Participants consisted of 15 university students enrolled in a spacecraft operations senior capstone course. The average age of all participants is 23.8 years (*SD* = 4.0), including 3 females and 12 males. All participants underwent the same treatment consisting of a 10-minute computer-based pre-training session (simulating spacecraft ground control) followed by a 10-minute GBVR training session (simulating an on-orbit spacewalk) and post-test surveys (see Figure 1). Participants were immersed in the VR environment using a software package titled *Mission ISS* [5] and worked from a seated position to minimize simulator sickness [6]. Participants were equipped with a *Valve Index VR kit* consisting of two hand-held controllers and a head-mounted display [7]. The independent variable is GBVR training, and the final survey scores serve as the dependent variable. Survey results were captured using four validated scales outlined in section 2.1.

**Figure 1. Spacecraft Operations Laboratories: Computer-Based Pre-training Simulation (left) GBVR Simulation (right)**

### *2.1. Validated Scales*

Several previously validated scales were employed during this study to evaluate whether the instructional design of course material and laboratory tools met specific criteria. The list of criteria includes adequate system usability, appropriate user satisfaction, balanced cognitive loading, and minimal simulator sickness. The following subsections will describe the scales used to measure each attribute.

#### *2.1.1. System Usability*

The *System Usability Scale* (SUS) was developed in 1986 as a subjective assessment tool for evaluating user perception of hardware devices and software applications regarding system complexity, ease of use, functionality, and user confidence [8]. The SUS survey will be used to evaluate the laboratory setup and equipment. The SUS survey contains ten questions rated on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). Final composite scores are ranked on a scale from 0 to 100 (0 = Worst Imaginable, 100 = Best Imaginable) [9], with a score of $M = 68$ being the published average standard [8].

### 2.1.2. User Satisfaction

The *Game User Experience Satisfaction Scale* (GUESS) [10] was developed in 2016 as a 55-question survey to measure user satisfaction and enjoyment during gameplay and later revalidated as an 18-question survey (GUESS-18) [11] to be used for this study. The questions are rated on a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree) covering nine constructs: usability, narratives, play engrossment, enjoyment, creative freedom, audio aesthetics, personal gratification, social connectivity, and visual aesthetics [11]. Scores are tabulated by summing the averages across all nine subscales and dividing by the maximum score of 63, resulting in a final score ranging from 0 (worst) to 100 (best). Six popular video games tested with GUESS-24 produced an average score of $M = 78.7$ (49.6 raw score divided by 63) [12] and will be used as the standard for this study. Survey wording will be modified from “play/playing” and “game” to “operate/operating” and “sim,” respectively.

### 2.1.3. Cognitive Loading

The *Cognitive Load Survey* (CLS) was developed in 2013 to measure the interactions between the various types of mental loading imposed on the learner during exposure to instructional material, tools, and strategies [13]. The CLS survey is comprised of ten questions measuring the interaction between the various types of cognitive loading and rated on a scale from 0 (not the case at all) to 10 (completely the case). Final scores are averaged for each loading type and ranked on a scale from 0 (low) to 10 (high). According to cognitive load theory [14], there are three types of cognitive loading: intrinsic, extraneous, and germane. Intrinsic cognitive load (IL) relates to task complexity and should be kept in a medium to low range (Approx. 2-5) to avoid disengagement of the learner due to tasks being either overly complex or exceptionally easy. Extraneous cognitive load (EL) impedes the learning process due to nonessential instructional elements and should be kept to a minimal level (Approx. 0-2). On the contrary, germane cognitive load (GL) refers to instructional features beneficial to learning and should fall within the medium to high range (Approx. 5-10). Furthermore, the *Multidimensional Cognitive Load Scale for Virtual Environments* (MCLSVE) [15] was developed later in 2018, adding four EL questions to the original survey regarding virtual environments (EL_vr), and will be used for this study.

### 2.1.4. Simulator Sickness

The *Simulator Sickness Questionnaire* (SSQ) was developed in 1993 to measure simulator-induced symptoms of nausea, oculomotor eye strain, and disorientation [16]. The SSQ survey consists of 16 questions rating each symptom on a 4-point Likert scale (0=none, 1=slight, 2=moderate, and 3=severe). Later in 2018, the *Virtual Reality Sickness Questionnaire* (VRSQ) was derived from the SSQ by reducing the survey to 9 questions [17]. VRSQ researchers eliminated the 7-question nausea category due to low reporting of nausea symptoms during their research trials with VR applications [17]. Consequently, the VRSQ will be used for this study. Final composite scores are tabulated based on proportional weighting of each symptom and rated on a scale of 0 (no symptoms) to 100 (severest symptoms). Studies show that longer immersion time will likely increase self-reported post-test symptom severity [18]. Since the average time of participant exposure for this study was approximately 10 minutes, the 0-15 minute range will be used, indicating an average symptom severity score of less than $M = 9.5$ [18].

## 3. Results

One-sample t-tests were conducted to compare participant results to the benchmark standards for the SUS and GUESS-18 measurement scales (see Table 1, Figure 2, and Figure 3). For the SUS survey data, the results indicated significantly higher scores for the simulation group ($M = 88.2$) compared to the accepted average score ($M = 68.0$), $t(14) = 5.88$, $p < .001$. A large effect size of $d = 1.52$ was revealed, demonstrating that the participants found the simulation relatively easy to use. For the GUESS-18 survey data, the results indicated significantly higher scores for the simulation group survey score ($M = 86.7$) compared to the average popular game score ($M = 78.7$), $t(14) = 3.87$, $p < .001$. A large effect size of $d = 0.99$ was indicated, signifying a high level of user enjoyment and satisfaction.

The mean comparisons for the MCLSVE and VRSQ results to the benchmark standards can be found in Table 1, Figure 4, and Figure 5. The mean MCLSVE scores ($M_{IL}$ = 2.4, $M_{EL}$ = 1.6, $M_{EL\_vr}$ = 1.4, and $M_{GL}$ = 8.7) placed within the approximate accepted ranges (0 < $M_{IL}$ < 2, 2 < $M_{EL}$ < 5, 2 < $M_{EL\_vr}$ < 5, and 5 < $M_{GL}$ < 10) [13]. Finally, the average VRSQ scores ($M_{Avg}$ = 6.2, $M_{Dis}$ = 5.8, $M_{Ocu}$ = 6.7) also placed within the accepted range (0 < $M$ < 9.5) [18].

**Table 1. Study Results vs. Standard Benchmarks [8] [11] [13] [18]**

| | *N* | Min | Max | *M* | *SD* | Standard (*M*) |
|---|---|---|---|---|---|---|
| **SUS** | 15 | 50.0 | 100.0 | 88.2 | 13.3 | 68.0 |
| **GUESS-18** | 15 | 73.0 | 99.2 | 86.7 | 8.0 | 78.7 |
| **MCLSVE IL** | 15 | 1.0 | 5.3 | 2.4 | 1.5 | Approx. 2-5 |
| **MCLSVE EL** | 15 | 1.0 | 4.7 | 1.6 | 1.0 | Approx. 0-2 |
| **MCLSVE ELvr** | 15 | 1.0 | 4.0 | 1.4 | 0.8 | Approx. 0-2 |
| **MCLSVE GL** | 15 | 6.5 | 10.0 | 8.7 | 1.2 | Approx. 5-10 |
| **VRSQ** | 15 | 0 | 25.8 | 6.2 | 8.0 | 0-9.5 |

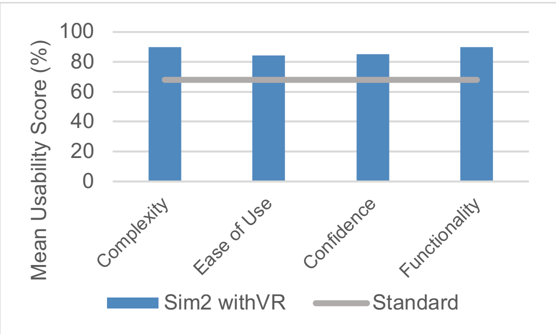


**Figure 2. System Usability Scale (SUS) [8]**

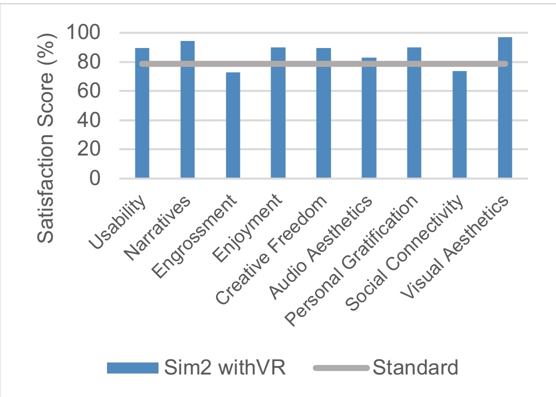


**Figure 3. Game User Experience Satisfaction Scale (GUESS-18) [11]**

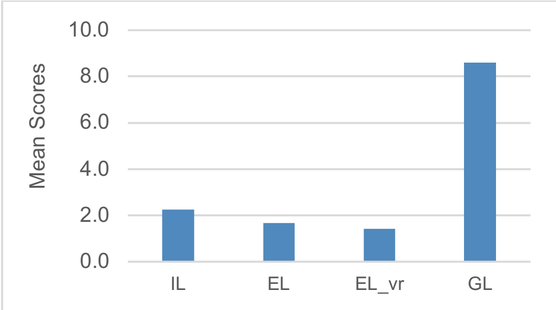


**Figure 4. Multidimensional Cognitive Load Scale for Virtual Environments (MCLSVE) [15]**

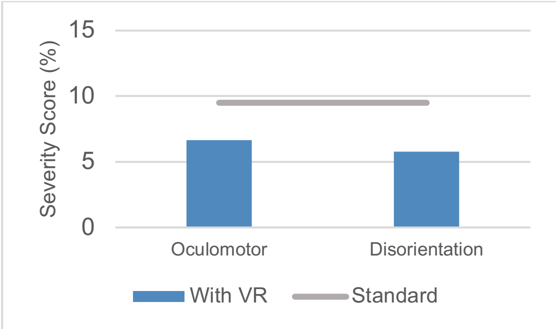


**Figure 5. Virtual Reality Sickness Questionnaire (VRSQ) [17] [18]**

## 4. Discussion

Compared to accepted benchmark standards, the GBVR instructional design employed in this study demonstrated satisfactory results in all categories, including system usability, user satisfaction, cognitive loading, and simulator sickness. As shown in Figure 2, system usability scores rated excellent along the SUS adjective scale [9] in complexity, ease of use, user confidence, and functionality. Successful usability scores are likely due to the effective laboratory setup, including the virtual reality simulation software (*Mission ISS* by Magnopus, 2019), game controllers, and head-mounted displays (*Valve Index* by Valve Corp., 2022).

Overall GUESS-18 user satisfaction scores ($M$ = 86.7) in Table 1 scored significantly higher than the benchmark standard ($M$ = 78.7). As shown in Figure 3, high scores in enjoyment and personal gratification indicate user motivation and interest in completing tasks skillfully, which are fundamental to student learning. The high level of user satisfaction is likely due to effective game mechanics such as game narrative, aesthetics, and goal accomplishment, including immediate feedback and reward. All categories exceeded the average popular game score ($M$ = 78.7) except for social connectivity ($M$ = 73.8) and play engrossment ($M$ = 72.9) (see Figure

3). This is likely due to the single-player educational activity offering no in-game social connection, like in the case of mainstream gaming communities. Conversely, during gameplay, the instructor gave verbal direction from outside of the GBVR environment. This interaction with someone outside the game may have slightly deterred play engrossment.

Notably, the SUS and GUESS-18 scales illustrate convergent validity regarding system usability. The overall SUS usability score ($M$ = 88.2) and the GUESS-18 usability subscale score ($M$ = 89.5) differ by only 1.3%, depicting converging scales (see Table 1, Figure 2, and Figure 3). This similarity further validates that survey questions from both scales accurately capture participant perceptions of system usability along with neighboring constructs of each scale.

The MCLSVE results displayed in Table 1 and Figure 4 indicate that appropriate cognitive balancing was imposed on the participants. Intrinsic loading ranked properly above 2.0 ($M_{IL}$ = 2.4), while extraneous loading ranked appropriately below 2.0 ($M_{EL}$ = 1.6, $M_{EL\_vr}$ = 1.4). Based on the accurate balancing of IL and EL, the remaining availability of participants' mental processing capacity contributed to high levels of germane loading, above 5.0 ($M_{GL}$ = 8.7). Successful cognitive balancing is likely due to the proper instructional design of the curriculum content. Task complexity adequately matched the learner's skill level, while nonessential extraneous loading was kept to a minimum.

Lastly, the VRSQ scores indicate low severity of symptoms due to VR simulation (see Table 1 and Figure 5). The average disorientation score ($M_{Dis}$ = 5.8) ranked well below the known average limit ($M_{Avg}$ = 9.5) [18]. Likewise, the average oculomotor score ($M_{Ocu}$ = 6.7) ranked well below the known average limit ($M_{Avg}$ = 9.5) [18]. These results are likely attributed to effective lab equipment and students performing VR activities from a seated position [6].

A limitation of this study was the small sample size ($n$ = 15), as this could restrict generalizability over the target population. Although the sample was small, the results were significant, and large effects were generated within the group of participants. However, repeating the study with a larger sample could further improve generalizability and external validity.

## 5. Conclusions

The integration and evaluation of GBVR in the classroom revealed noteworthy results. First, based on proper laboratory setup, system usability rated excellent along the SUS adjective scale [9] in complexity, ease of use, user confidence, and functionality. Second, based on effective game mechanics, overall user satisfaction ranked significantly higher than six popular video games analyzed by Shelstad et al. (2019). Third, cognitive loading was adequately balanced based on proper instructional design, facilitating student learning. Lastly, simulator sickness did not exceed acceptable minimums due to an effective laboratory format. These positive results set the foundation for potentially enhanced student learning. With GBVR correctly integrated into the classroom or training environment, learner enjoyment and satisfaction may be amplified, likely leading to increased motivation, cognitive engagement, and skill retention [1] [4].

## Acknowledgements

I recognize all student participants for their contributions and excitement for spacecraft operations.

Thank you, Trevor Goodwin and the XR lab team, for their support and assistance with VR laboratory equipment and activities.

Thank you, Embry Riddle Aeronautical University, College of Aviation leadership for the development and continued support of the XR laboratory facilities available to all faculty and students.

Thank you, Dr. Stephanie Fussell, for sharing your expert knowledge, guidance, and research in education and training with virtual reality applications.

## References

[1] R. M. Ryan, E. L. Deci, Self-determination theory: basic psychological needs in motivation, development, and wellness. Guildford Publications, 2018.

[2] Y. Zhonggen, A meta-analysis of use of serious games in education over a decade, *International Journal of Computer Games Technology*, 2019.

[3] A. Christopoulos, N. Pellas, and M. Laakso, A learning analytics theoretical

framework for STEM education virtual reality applications, *Education Sciences*, 10(317), 317, 2020.

[4] M. Csikszentmihalyi, and K. Asakawa, Universal and cultural dimensions of optimal experiences: Flow, culture, and human evolution, *Japanese Psychological Research*, 58(1), 4-13, 2016.

[5] Magnopus Mission ISS Website: https://missioniss.magnopus.com, last visited 2nd March 2022.

[6] Y. Hu, M. Elwardy, and H. Zepernick, On the effect of standing and seated viewing of 360◦ videos on subjective quality assessment: A pilot study, *Computers* (Basel), 10(6), 80, 2021.

[7] Valve Corporation Website: https://www.valvesoftware.com/en/index, last visited 2nd March 2022.

[8] Usability.gov Website: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html, last visited: 2nd March 2022.

[9] A. Bangor, P. T. Kortum, and J. T. Miller, Determining what individual SUS scores mean: Adding an adjective rating scale, *Journal of Usability Studies*, 4(3), 114-123, 2009.

[10] M. H. Phan, J. R. Keebler, and B. S. Chaparro, The development and validation of the game user experience satisfaction scale (GUESS), *Human Factors*, 58(8), 1217-1247, 2016.

[11] J. R. Keebler, W. J. Shelstad, D. C. Smith, B. S. Chaparro, and M. H. Phan, Validation of the GUESS-18: A short version of the game user experience satisfaction scale (GUESS), *Journal of Usability Studies*, 16(1), 49 – 62, 2020.

[12] W. J. Shelstad, B. S. Chaparro, and J. R. Keebler, 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, 2019.

[13] J. Leppink, F. Paas, C. P. M. v. d. Vleuten, T. A. J. M. v. Gog, and J. J. V. Merrienboer, Development of an instrument for measuring different types of cognitive load, *Behavior Research Methods*, 45(4), 1058-1072, 2013.

[14] J. Sweller, Element interactivity and intrinsic, extraneous, and germane cognitive load, *Educational Psychology Review*, 22(2), 123-138, 2010.

[15] M. S. Andersen, and G. Makransky, The validation and further development of a multidimensional cognitive load scale for virtual environments, *Journal of Computer Assisted Learning*, 37(1), 183-196, 2020.

[16] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness, *The International Journal of Aviation Psychology*, 3:3, 203-220, 1993.

[17] K. H. Kim, J. Park, Y. Choi, and M. Choe, Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment, *Applied Ergonomics, 69*, 66-73, 2018.

[18] B. K. Jaeger, and R. R. Mourant, Comparison of simulator sickness using static and dynamic walking simulators, *Human Factors and Ergonomics Society Annual Meeting*, 45(27), 1896-1900, Minnesota, USA, 2001.