# CS 549 HUMAN COMPUTER INTERACTION
# ASSIGNMENT 5

# END-USER BASED USABILITY TESTING OF VIRTUAL REALITY ENVIRONMENTS


**Group18**

Ahmet Yasin Aytar

Atra Zeynep Bahçeci

Halil İbrahim Ergül

Serhat Demirkıran


26 December 2023

## 1. Methodology

This study is performed to investigate the user preferences between 2 roller coaster VR (Virtual Reality) applications CoSpaces and YouTube Jurassic Dinosaur Coaster, and to identify the factors related to VR immersion.

User-based usability testing is done with 4 end users. The user persona is created according to the target audience of the applications and the actual demographics of VR users. 57% of VR users are male, between the ages of 25-34 (Williams, 2023). So the target demographic is young, not dominated by a single gender, and also “tech-savvy”. The persona is given below:

**Name:** Elif Kaya

**Occupation:** Master's Student in Computer Science

**Education:** BSc in Computer Engineering, Currently pursuing MSc in Computer Science

**Age:** 25.

**Profile:** Elif is a dedicated master's student at Sabanci University, Turkey. Her academic success in computer science is fueled by her attention to detail and a strong interest in emerging technologies. Fluent in Turkish and English, she thrives in an academic environment where innovation and tech are at the forefront.

**Goals:**

- To gain practical experience in the field of virtual reality, which she believes is the future of user experience.
- To contribute to the development of user-friendly and immersive VR applications.
- To enhance her knowledge in user-interface design and user-experience research.

**Needs:**

- Intuitive and engaging VR applications that can be used as learning tools ● Access to technology that can seamlessly blend with her academic projects.
- A platform to share her research and insights in the tech community.

**Technical Background and Workplace:** With a solid foundation in computer engineering and a growing expertise in computer science, Elif is adept with technology. She spends a considerable amount of time on her MacBook Pro, coding and designing software. Her tech-savviness is evident from her preference for Apple products – she owns an iPhone 12 and uses an iPad for reading and research. Her technical skills are not just limited to her field of study; she is equally passionate about understanding user behavior and the psychology behind it.

**General Background:** Elif is a curious and meticulous individual, always keen to explore the latest tech trends. She has never used VR before, making her an ideal candidate for testing the immersive aspect of VR applications without preconceived notions. She has a balanced lifestyle, juggling academic excellence with hobbies like reading and exploring the tech world through blogs and podcasts. Her favorite brands include Apple and Google, reflecting her taste for quality and innovation.

**Archetype:** Elif represents the archetype of a 'Tech Enthusiast' and 'Diligent Student' – she is methodical in her approach to new technologies and believes in leveraging them for enhancing academic and professional productivity.

**Experience Goals:** Elif expects VR applications to be not just entertaining but also informative. She looks forward to experiencing a level of immersion that could potentially be used in educational settings, making learning an enjoyable venture.

**Brand-Relationship:** She is loyal to brands that consistently deliver innovation and user-friendly experiences. Apple's ecosystem has won her over due to its seamless integration and intuitive design.

*User Persona*

Before conducting the test, the demographic information of participants is taken via a questionnaire form (link given in the Appendix as the *Pre-Test Questionnaire)*. A "Prefer not to say" option is added to all the questions which may be sensitive to some users, and the ages are collected via an interval selection as opposed to direct entry, as suggested by Tankala (2022). In addition to the demographics, participants' existing perceptions of VR and roller coasters are also recorded via this questionnaire. The results are given below in *Table 1*.

| **Participant ID** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| **Gender** | Female | Male | Male | Female |
| **Age** | 20-29 | 20-29 | 20-29 | 20-29 |
| **Education Level** | Master's | Undergraduate | Master's | Master's |
| **Occupation** | NLP Engineer | Student | Student | Computer Engineer |
| **Nationality** | Turkish | Turkish | Syrian | Turkish |
| **Residency** | Istanbul | Istanbul | Istanbul | Istanbul |
| **Had previous VR experience** | No | No | No | No |
| **Had previous experience with roller coasters** | No | Yes | Yes | Yes |
| **Overall Computer Skills (1-5)** | 4 | 4 | 5 | 5 |
| **Opinion on VR Technologies (1-5)** | 5 | 5 | 4 | 5 |
| **Interest in VR Technologies (1-5)** | 5 | 3 | 2 | 3 |
| **Level of Comfort with Android Technologies (1-5)** | 5 | 5 | 4 | 4 |
| **Wearing Glasses** | No | Yes | No | Yes |
| **Perceived Visual Ability (1-5)** | 4 | 5 | 2 | 3 |
| **Perceived Hearing Ability (1-5)** | 5 | 5 | 4 | 5 |
| **Has physical / cognitive impairments** | No | No | No | No |

*Table 1: Results of the Pre-Test Questionnaire*

The context is the user watching the VR videos in a quiet, calm environment; sitting down in a rotating chair. The tool is the card box VR glasses and an Android phone with dimensions 159.3 × 74 mm. Since the applications studied are not interactive, the task given to the users is to simply experience the 2 virtual environments.

Before the test, the participants were informed that they would use 2 applications in which they would be on a roller coaster ride. They were also informed that the applications are not interactive and that they will use each application for about 120 seconds. Before each application, users were allowed to "get the feel" of each application, look around in the virtual environment, get adjusted to the vision, etc. A photo of each user during the test is shown in *Figure 1*. Please note that as the dimensions of the phone used were bigger than the slot allotted

for phones in the card box glasses, users had to hold up the glasses themselves. 2 of the users started with the YouTube Jurassic Dinosaur Coaster, and 2 started with the CoSpaces application. The users tested the applications back-to-back, with a few minutes in between them depending on the participant's preferences. Once the test was completed, the participants completed 2 additional questionnaires (links given as the *Post-Test Interview* and the *System Usability Scale* in the Appendix). The Post-Test Interview questions were directed toward measuring the user preferences between the 2 applications, the immersion success of the 2 applications, and also the factors that strengthen/inhibit immersion. Some of the questions were adapted from a previous study that specifically investigated VR immersion by Tcha-Tokey (2016). The SUS (System Usability Scale) questionnaire allowed for quantitative data relating to both applications. The participants were directed to complete the SUS form before any discussion, and to not dwell on their answers as suggested by Brooke (1996).

## 2. Results

During the VR usability test, participants displayed a range of reactions, indicative of their immersion and the applications' ability to take responses. Participants 1 and 3, who both started with CoSpaces, showed physical engagement: Participant 1 with laughter and surprise, indicative of a playful and engaging experience, and Participant 3 through rotational movements, though his engagement seemed more moderate as he did not look up or down. Participants 2 and 4, who began with the YouTube VR, also demonstrated physical reactions; Participant 2's leg twitch and disorientation—losing his sense of direction and ending up standing backwards—suggest a deeply immersive experience that perhaps lacked spatial anchors, while Participant 4's laughter and head movements signaled enjoyment and active exploration.

Post-test comments further differentiated the two experiences: CoSpaces was described as visually tiring and more intense with its spins and rotations, potentially contributing to feelings of fatigue and motion sickness in some users. In contrast, the YouTube VR was noted for clearer vision and smoother transitions, though it still caused some disorientation and physical reactions such as shoulder tension and considerable head tilting. These observations highlight the distinct immersive qualities and potential sensory impacts of each VR application, with CoSpaces revealing a more physically active response, while the YouTube experience seemed to be smoother but not without its own set of challenges.

A post-interview survey was administered to four participants after they had experienced VR through VR glasses. The data collected from this survey as shown in *Table 2* reveals insights into various aspects of the participants' experiences with two different VR applications, namely a 3D YouTube video and a 3D application called CoSpaces. The survey results indicate that cyber-sickness was experienced by three of the four participants, with only one participant reporting no such symptoms. Despite this, all participants expressed a

preference for the YouTube application. In terms of comfort while using the VR headset, the responses varied: one participant felt moderately comfortable, another slightly more so, and a third less comfortable, suggesting a range of user comfort levels with the VR technology. As *Figure 2* demonstrates, the immersion scores given to CoSpaces averaged out to moderate compared to Youtube, with one participant rating it highly while another found it less immersive. The visual aspects of CoSpaces were rated fairly positively, although there was one standout high rating. Audio scores for CoSpaces indicated a general satisfaction, with one participant marking it above the others. YouTube's immersion scores were high across the board, indicating a strong sense of presence within the VR environment. The visuals for YouTube were rated highly by one participant in particular, with the others providing moderately high scores. Audio quality in YouTube also received high marks, particularly from one participant who rated it significantly better than the others. Interestingly, all participants identified the YouTube application as the point at which cyber-sickness symptoms began to emerge. These findings reflect individual experiences and are indicative of varied user responses to the immersive qualities, audio-visual elements, and comfort levels associated with VR applications.

| **Participant ID** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|
| **Cyber-Sickness** | Yes | No | Yes | Yes |
| **Preferred Application** | YouTube | YouTube | YouTube | YouTube |
| **Comfort with VR Headset** | 3 | 3 | 4 | 2 |
| **CoSpaces Immersion Score** | 4.33 | 2.33 | 1.66 | 2.33 |
| **CoSpace Visuals Score** | 5 | 4.33 | 3.66 | 3.66 |
| **CoSpace Audio Score** | 3.66 | 4 | 3.33 | 3.66 |
| **YouTube Immersion Score** | 5 | 3.33 | 4 | 4 |
| **YouTube Visuals Score** | 5 | 3.33 | 3.66 | 4.66 |
| **YouTube Audio Score** | 3.66 | 3.33 | 3.33 | 4 |
| **Cyber-Sickness Onset Application** | YouTube | YouTube | YouTube | YouTube |

*Table 2: Post-Test Interview Results*

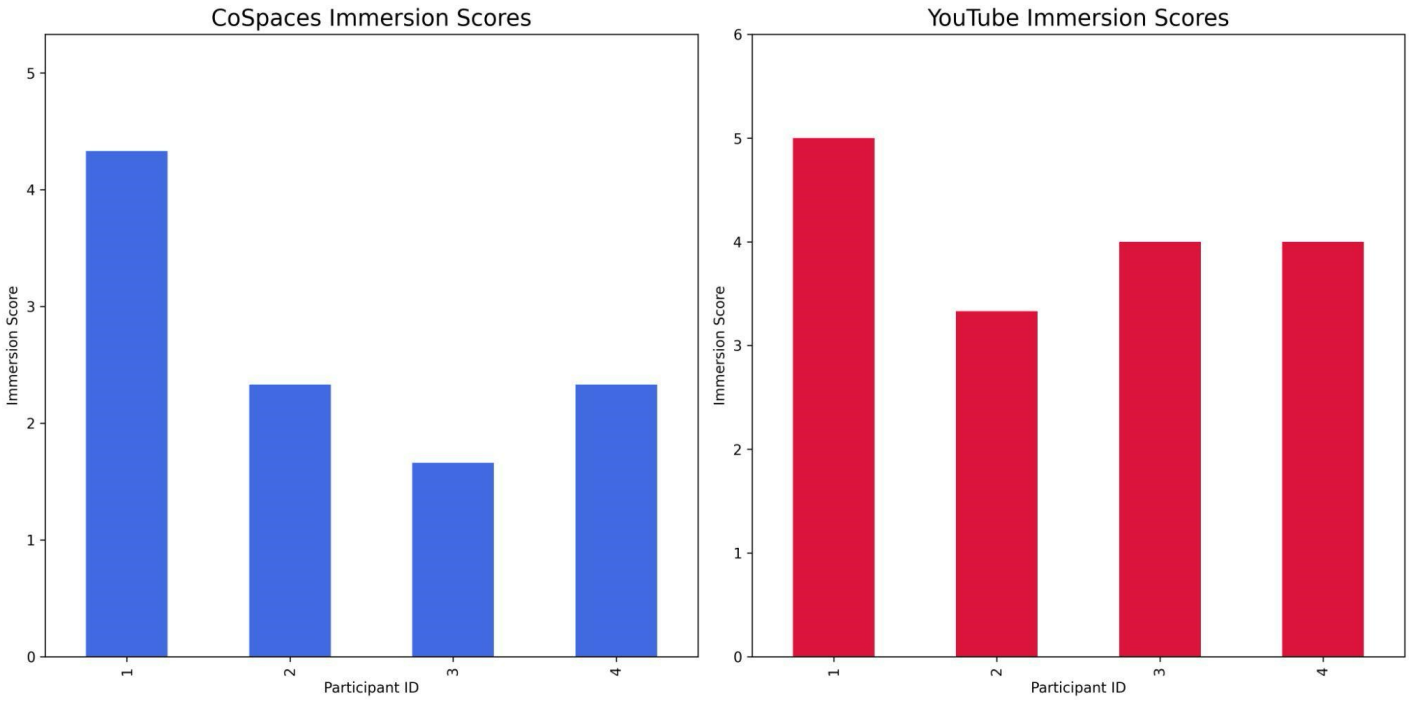


*Figure 2: Cospaces vs. Youtube Immersion Scores*

The post-assessment survey, as shown in *Table 3* for both CoSpaces and YouTube experiments, contains various metrics. The Ease of Use Score for CoSpaces ranged from 3 to 4.5, indicating a moderate level of usability with one participant finding it notably easier to use than the others. The Complexity Score for CoSpaces was predominantly rated as 1, suggesting it was not complex, although one participant rated it slightly higher at 1.5. The Learning Curve Score for CoSpaces was consistently rated as 3 across all participants, implying a moderate learning curve. Confidence Scores were high for CoSpaces, with two participants rating it as 5, indicating high confidence, and the other two slightly less so at 4. The Consistency Score for CoSpaces showed some variation, with one participant noting perfect consistency at a score of 1, while the others found it less consistent at a score of 2.

The SUS Score, a measure of overall usability, showed a high variance ranging from 65 to 95 for CoSpaces, reflecting significant differences in perceived usability among participants.

| Participant ID | Application | Ease of Use Score | Complexity Score | Learning Curve Score | Confidence Score | Consistency Score | SUS Score |
|---|---|---|---|---|---|---|---|
| 1 | CoSpace | 3.5 | 1 | 3 | 5 | 2 | 87.5 |
| 2 | CoSpace | 4.5 | 1 | 3 | 5 | 1 | 95 |
| 3 | CoSpace | 3.5 | 1 | 3 | 4 | 2 | 75 |
| 4 | CoSpace | 3 | 1.5 | 3 | 4 | 2 | 65 |
| 1 | Youtube | 3.5 | 1 | 3 | 5 | 1 | 92.5 |
| 2 | Youtube | 4 | 1.5 | 2.5 | 4 | 2 | 82.5 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3 | Youtube | 4 | 1 | 3 | 4 | 1 | 85 |
| 4 | Youtube | 3 | 2 | 3 | 4 | 2 | 67.5 |

*Table 3: Post-Test Assessment Results*

When assessing the YouTube application, the Ease of Use Score was similar to that of CoSpaces, with scores ranging from 3 to 4, again suggesting moderate usability. For Complexity, YouTube received varied scores, with two participants finding it not complex at a score of 1, one at a slightly more complex score of 1.5, and one participant rating it as more complex with a score of 2. The Learning Curve Score for YouTube was slightly lower than CoSpaces, indicating a somewhat easier learning process, particularly noted by one participant who rated it at 2.5. Confidence in using YouTube was rated at 4 or 5, showing that participants generally felt confident using the application. The Consistency Score for YouTube also displayed variation, with half the participants noting perfect consistency at a score of 1, and the other half noting less consistency at a score of 2. The SUS Score for YouTube varied from 67.5 to 92.5, pointing to different levels of satisfaction with the usability of the YouTube VR experience among the participants.

These post-assessment results provide a nuanced view of the participants' perceptions of the ease of use, complexity, learning curve, confidence, consistency, and overall usability (SUS Score) of the CoSpaces and YouTube applications as experienced through VR glasses. The scores highlight individual differences in how each participant interacted with and perceived the two applications.

## 3. Discussion and Interpretation

In the context of human-computer interaction (HCI), the study's findings offer insightful data on user engagement, usability, and the physiological responses to VR applications. This discussion will analyze and interpret the gathered data to provide a detailed account of user experience and interaction with the VR applications under consideration: a 3D YouTube video and the Cospaces application. In a nutshell, we will focus on what data tells us about the applications and how we can improve each application.

Several areas can be highlighted in interpreting and discussing our findings. To focus on user engagement and physical response first, one can say that our participants exhibited varied physical responses during the VR usability test, which are indicative of the immersive qualities of the VR applications. CoSpaces elicited more physical engagement, as seen in Participant 1's laughter and Participant 3's rotational movements. This suggests that CoSpaces may offer a more interactive and engaging experience. Conversely, the reactions to YouTube VR, such as Participant 2's disorientation and Participant 4's head movements, indicate a deep immersion that, while compelling, may lack "spatial anchors" (meaning their real-world location in coordinate system), leading to confusion. Secondly, despite some reporting cyber-

sickness, all participants expressed a preference for the YouTube VR application. This preference could be attributed to its clearer vision, smoother transitions, and a more comfortable and consistent user experience, as reflected in the survey scores.

What is more, *The Ease of Use* and *Learning Curve* scores suggest that both applications are moderately user-friendly. However, YouTube VR appears to have a slightly easier learning curve. This difference could be pivotal in first-time user experiences, where ease of adaptation can significantly impact user satisfaction. Cospaces received lower scores in several areas, including ease of use and consistency. This highlights the importance of usability in HCI, particularly for new applications. If a technology is perceived as complex or challenging to learn, it can lead to user frustration, which is evidenced by the lower confidence and consistency scores for Cospaces. In the HCI context, this suggests that a balance must be struck between offering advanced functionalities and maintaining user-friendly interfaces. Also, the visual aspects of both applications were rated positively, but YouTube VR seems to have an edge in this regard. The high audio scores for both applications indicate satisfactory acoustic performance, an essential aspect of immersive VR experiences.

Another part worth discussing with respect to our findings is cyber-sickness. It is an adverse reaction related to motion sickness that can occur in VR environments, and it was reported by three participants. This suggests that despite the high interest and computer skills, physiological responses can significantly impact user experience. From an HCI perspective, minimizing such negative responses is critical, as they can discourage users from engaging with the technology in the future. The absence of cyber-sickness in participant 2 might suggest individual differences in susceptibility.

Setting our findings aside, we can highlight briefly on how we can improve each application. For CoSpaces, enhancing user engagement could be informed by Bowman and McMahan's (2007) discussion on the importance of interaction integrity in virtual environments. This research suggests that more interactive and customizable features could lead to a more engaging user experience. Addressing the issues of visual fatigue and motion sickness in CoSpaces may benefit from the principles laid out by LaViola Jr. (2000), who discusses methods to mitigate VR-induced symptoms, such as optimizing motion dynamics and providing adjustable settings for motion sensitivity.

In the case of YouTube VR, the smoother transitions and clearer vision need to be maintained, while addressing the disorientation issues could be guided by the work of Ruddle et al. (1999). Their research indicates that clear spatial orientation cues, both visual and auditory, can help maintain user orientation and reduce disorientation, thereby enhancing the overall VR experience. Furthermore, the cyber-sickness challenges present in both applications can be approached by drawing upon Kennedy et al.'s (1993) research. Their work on simulator sickness in virtual environments provides valuable insights into how individual differences in

susceptibility can be addressed, suggesting solutions like adjusting the field of view and stabilizing camera movements to improve user comfort. By integrating these research-backed strategies into the development of CoSpaces and YouTube VR, the applications can be refined to enhance user engagement, reduce discomfort, and provide a more inclusive and accessible VR experience.

Lastly, this study is far from perfect due to its limitations. The limited sample size of our study, consisting of only four participants, is a significant constraint. The small cohort restricts the generalizability of our findings, as it does not adequately represent the diverse spectrum of potential users. A broader demographic representation is crucial for ensuring that the insights gained are applicable to a wider population. As highlighted in research by Nielsen and Landauer (1993), the number of test users plays a critical role in uncovering usability problems. A larger sample size would not only enhance the reliability of the findings but also provide a more nuanced understanding of varied user experiences and preferences.

The short duration and limited scope of exposure to the VR applications present another limitation. Participants interacted with the applications for a brief period, which limits our understanding of user adaptation and long-term engagement with the technology. Longitudinal studies, as discussed by Karahanna et al. (1999), are essential to understand how users adapt to and accept new technologies over time. Last but not least, our study's reliance on subjective measures, primarily self-reported data, introduces potential biases and does not capture all dimensions of user interaction and satisfaction. Self-reports can be influenced by various factors, including participants' mood, memory, and desire to give socially desirable responses. Objective measures, such as physiological data or task completion time, as discussed by Brooke (1996) in the development of the System Usability Scale (SUS), could complement self-reports to provide a more rounded understanding of the user experience. Incorporating a mix of subjective and objective measures can help in gaining a comprehensive understanding of the user interaction with VR applications.

## 4. Conclusion

This study conducted a user-based usability testing in VR environments. The data reveals distinct user preferences between the two VR applications, CoSpaces and YouTube Jurassic Dinosaur Coaster. Participants generally preferred the Youtube VR for its clearer vision and smoother transitions as evidenced by their higher immersion and comfort scores.
However, CoSpaces, despite causing some cyber-sickness and being visually tiring, was noted for its engaging and intense experience

The SUS scores reflect the usability and user satisfaction with two VR applications. Calculated scores for CoSpaces application indicates more diverse user experience. On the

other hand, scores for Youtube VR applications show more uniform and slightly higher user satisfaction. Participants noted a more comfortable and less disorienting experience for Youtube VR, explaining its overall positive but less varied scores compared to CoSpaces.

Future studies in the field of VR should prioritize increasing the sample size and diversity. A larger and more varied participant pool would not only validate the findings but also ensure they are reflective of a broad spectrum of user experiences. This is crucial for understanding how different groups, with their unique preferences and needs, interact with and perceive VR applications. A study with a diverse participant base would provide insights that are more representative of the general population, making the findings more applicable to a wider audience.

In addition to expanding the participant demographic, adopting a longitudinal study design would offer valuable insights into long-term user engagement and adaptation to VR technologies. By observing how users interact with VR applications over an extended period, researchers can gain a deeper understanding of the evolving nature of user experiences. This approach is particularly useful in identifying trends in user behavior, preferences, and potential issues that may arise only after repeated use or prolonged exposure to the technology.

## Appendix

- Pre-Test Questionnaire
- Post-Test Interview
- System Usability Scale

## References


1. Bowman, D., & McMahan, R. (2007). Virtual Reality: How Much Immersion Is Enough? *Computer*, *40(7)*, 36–43. https://doi.org/10.1109/MC.2007.257
2. Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, I. L. McCelland, & B. Weerdmeester (Eds.), *Usability Evaluation in Industry* (1st ed., pp. 189-194). CRC Press. https://doi.org/10.1201/9781498710411
3. Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information Technology Adoption Across Time: A Cross-Sectional Comparison of Pre-Adoption and Post-Adoption Beliefs. *MIS Quarterly*, *23*(2), 183–213. https://doi.org/10.2307/249751
4. LaViola, J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, *32(1)*, 47–56. https://doi.org/10.1145/333329.333344
5. Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. *Proceedings of the INTERACT '93 and CHI '93 Conference on*

*Human Factors in Computing Systems*, 206–213. https://doi.org/10.1145/169059.169166

6. Ruddle, R., Payne, S., & Jones, D. (1997). Navigating Buildings in Desk-Top Virtual Environments: Experimental Investigations Using Extended Navigational Experience. *Journal of Experimental Psychology: Applied*, *3*, 143–159. https://doi.org/10.1037/1076-898X.3.2.143
7. Tankala, S. (2022, November 27). *Why and How to Use Demographics in UX*. Nielsen Norman Group. https://www.nngroup.com/articles/demographics-in-ux/
8. Tcha-Tokey, K., Christmann, O., Loup-Escande, E., & Richir, S. (2016). Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments. *The International Journal of Virtual Reality*, *16*, 33–48. https://doi.org/10.20870/IJVR.2016.16.1.2880
9. Williams, J. (2023, October 19). *Virtual Reality Demographics: 57 User Facts & Numbers*. Techpenny. https://techpenny.com/virtual-reality-demographics-57-user-facts-numbers-2023/