

# Virtual Reality for Training Diagnostic Skills in Anorexia Nervosa: A Usability Assessment

Jose Gutierrez-Maldonado<sup>(✉)</sup>, Antonio Andres-Pueyo, Adolfo Jarne, Antoni Talarn, Marta Ferrer, and Joseba Achotegui

Department of Clinical Psychology and Psychobiology,  
University of Barcelona, Barcelona, Spain  
jgutierrezm@ub.edu

**Abstract.** Virtual Reality (VR) technology is used in clinical psychology to integrate and enhance traditional assessment and therapeutic approaches for a variety of conditions. It is also increasingly used in the training of health professionals, as it provides authentic recreations of real-life settings, without exposing students to situations for which they are not yet prepared. VR systems involve different graphical user interfaces for human–computer interaction that vary according to the level of immersion required. In this study, we explore the interaction between level of immersion and gender, in order to establish whether the differences in usability between men and women found in previous studies are modulated by the level of immersion of the VR devices used to perform the simulations. Seventy undergraduate students (44 women, 26 men) participated in the study. They were randomly assigned to one of the two following conditions: differential diagnosis skills training using simulated interviews with an immersive system, or training using the simulated interviews with a non-immersive system. The results showed that men rated the usability of immersive and non-immersive systems to be almost the same, while women assessed the usability of the non-immersive system to be higher. A greater proneness to motion sickness in women is proposed as a hypothesis to explain these differences; this hypothesis should now be tested in further studies.

**Keywords:** Virtual reality · Training · Psychology · Usability · Immersion

## 1 Introduction

Virtual Reality (VR) technology is used in clinical psychology to integrate and enhance traditional assessment and therapeutic approaches for a variety of conditions. The first studies of its effectiveness in the treatment of psychological disorders concentrated on different types of phobia; these disorders are often approached via exposure therapies, for which VR is particularly well suited. In a recent publication [1], Riva et al. reported on the available reviews and meta-analyses on the use of VR in clinical and health psychology. VR exposure has been demonstrated to be efficacious for the treatment of a variety of psychological disorders, offering several advantages such as high ecological validity, high acceptability, and increased control over variables [2, 3].

Virtual reality environments are also increasingly used in the training of professionals, as they very effectively reproduce real-life settings without forcing students to deal with situations for which they are not yet prepared. VR can grab the attention in a far more immediate way than other kinds of media, drawing students inside educational experiences that cannot be carried out using other techniques [4].

VR systems involve different graphical user interfaces for human–computer interaction that vary according to the level of immersion required. The most basic level involves exposure to virtual environments on computer screens, with peripheral input devices (e.g., a keyboard or a computer mouse) used to interact with them. At the other extreme are innovative and technologically advanced systems such as Head Mounted Displays (HMD) which simulate binocularly overlapped images and create the illusion of a three-dimensional world. VR provides trainees with simulations of real-life situations in which they can learn by doing in a safe educational context, and allows trainers to gradually increase the difficulty of the problems to be solved in the training tasks; this facilitates the process of learning by guiding students towards their optimal performance. The implementation of VR-based applications for training has always depended heavily on the development of advanced technology, and so for a long time the development in this area was limited by the cost of the equipment required. However, this scenario is now changing, due to the expansion of VR in the field of consumer electronics; the commercialization of VR systems among the general population is bringing down costs and enhancing the development of user-friendly devices. Furthermore, for younger generations the use of VR technology will be part of their everyday routine and the technical difficulties will disappear [5].

In a previous study, we compared the usability of two low-cost VR systems which offered different levels of immersion for training students in diagnostic interview skills, by means of simulations of psychopathological examinations in patients with eating disorders [6]. Previous research has shown these simulations to be more effective for the training of differential diagnosis skills than traditional methods based on role-playing [7]. No differences were found in usability in our earlier study between immersive and non-immersive systems. Given the greater complexity and higher cost of immersive systems, it was concluded that non-immersive systems are a promising VR alternative for developing these skills in trainee professionals. In another study [8], in order to establish whether individual differences such as gender should be taken into account in the design of training systems of this kind, we compared the usability scores of male and female students, finding significantly higher usability scores for men on several items of the Software Usability Measurement Inventory (SUMI) [9].

In the present study, we explored the interaction between these two variables – level of immersion and gender – in order to establish whether the differences in usability between men and women are modulated by the level of immersion of the VR devices used to perform the simulated interviews.

## 2 Method

### 2.1 Participants

Seventy undergraduate students (44 women, 26 men) participated in the study. They were randomly assigned to one of the two following conditions: differential diagnosis skills training using the simulated interviews with an immersive system (Oculus Rift DK2), or training using the simulated interviews with a non-immersive system (a stereoscopic computer screen) (Fig. 1). A restriction to the random assignment was that each group had to have the same number of subjects (35); thus, if the previous participant was assigned randomly to one condition, the following participant was assigned to the other condition. One participant in the immersive condition was unable to complete the simulated interview because of motor sickness (she reported mild nausea); another participant in the non-immersive condition was not able to complete the simulated interview because of a malfunction of the computer that could not be satisfactorily repaired at that moment. Both participants who did not finish the experiment were women. Finally, 68 participants (42 women, 26 men) completed the study.



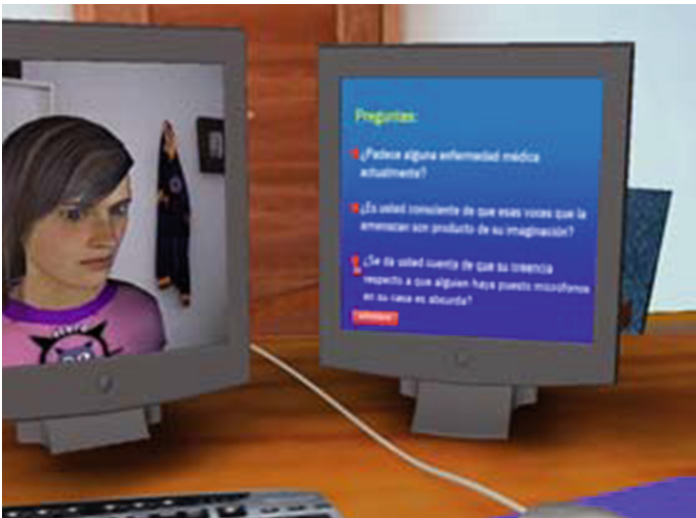
**Fig. 1.** Oculus Rift DK2 (above), and Acer Aspire 5738DG laptop with stereoscopic display (below)

## 2.2 Instruments

The non-immersive virtual interviews were displayed on an Intel Pentium T4400 IV (2.2 GHz, 800 MHz FSB) laptop with 4 GB RAM, ATI Mobility Radeon HD 4570 graphics card, and a 15.6 inch 3D monitor. Earphones and polarized glasses were used. The immersive virtual interviews were displayed on an Oculus Rift VR HMD DK2 system. The Oculus provided immersive 3D virtual environments in a wide field of vision (100°), OLED screens with a resolution of 960 × 1080 per eye with low head-tracking latency (20 ms) and high refresh rate (75 Hz). Earphones were also used.

## 2.3 Procedure

In the virtual simulations (Fig. 2), learners conducted a clinical interview with different Virtual Patients (VPs). Each VP presented a specific eating disorder. The objective of the interviews was to obtain enough data to formulate a diagnosis. To do so, users selected the most suitable question at each stage of the interview; the system informed them how accurate their choice was, and the VP responded to their questions. At each stage, users decided whether to continue asking questions or whether they had enough information to formulate a diagnostic hypothesis. If they selected the correct diagnosis at any given time during the interview, the system would only accept it if the VP had been fully examined. When the simulation had been completed, students were asked to evaluate the usability of the system. The mean duration of the virtual interview was 32 min (SD = 8 min).



**Fig. 2.** Virtual interview: The VP appears on the left-hand screen, while the question choices and the diagnosis hypothesis are displayed on the right-hand screen.

Usability was assessed with the Software Usability Measurement Inventory (SUMI). Only the items on the inventory that are applicable to our software were considered for data analysis:

2. I would recommend this software to my colleagues
3. The instructions and prompts are helpful
5. Learning to operate this software initially is full of problems
7. I enjoy my sessions with this software
12. Working with this software is satisfying
13. The way that system information is presented is clear and understandable
17. Working with this software is mentally stimulating
19. I feel in command of this software when I am using it
26. Tasks can be performed in a straightforward manner using this software
27. Using this software is frustrating
29. The speed of this software is fast enough
32. There have been times in using this software when I have felt quite tense
42. The software has a very attractive presentation
44. It is relatively easy to move from one part of a task to another
48. It is easy to see at a glance what the options are at each stage.

For each of the items on the SUMI, participants had to select one of three options: agree, undecided, or disagree. To obtain an overall usability score, on items 2, 3, 7, 12, 13, 17, 19, 26, 29, 42, 44, and 48, positive answers (agree) scored 1 point, “undecided” answers scored 0 points, and negative answers (disagree) were assigned a score of  $-1$ . On items 5, 27 and 32, on the other hand, negative answers (disagree) were assigned a score of 1, positive answers (agree) a score of  $-1$ , and “undecided” answers scored 0 points.

In order to analyse the influence of the level of immersion and gender on performance, students in both groups were required to take an anorexia nervosa diagnostic interview skills test comprising 50 written questions; the final score was calculated taking into account the correct answers converted to a 10-point scale.

### 3 Results

The scores on the overall measure of usability were similar in the immersive and the non-immersive systems (mean = 9.08, SD = 4.27 in the HMD group; and mean = 10.23, SD = 2.52 in the stereoscopic screen group;  $F = 1.57$ ,  $p = 0.21$ ). However, men gave higher usability scores than women (mean = 11.5, SD = 2.45 in men; and mean = 8.52, SD = 3.63 in women;  $F = 14.73$ ,  $p < 0.001$ ). The interaction between level of immersion and gender was very near to reaching significance ( $F = 3.4$ ;  $p = 0.07$ ). As can be seen in Table 1, the usability of immersive and non-immersive systems was almost the same for men, while women gave significantly lower ratings for the usability of the immersive system.

The influence of gender and level of immersion on performance was analysed by means of an ANOVA applied to a  $2 \times 2$  between-subjects design (women/men, immersive/non-immersive system). Neither the principal effect nor the interaction between factors was significant. The learning achievements of women and men after the VR

**Table 1.** Usability of immersive and non-immersive systems for women and men

Gender	Immersion	Mean usability	SD	N
Men	HMD (Immersive)	11.7143	2.33464	14
	Stereo screen (Non-immersive)	11.2500	2.66714	12
	Total	11.5000	2.45357	26
Women	HMD (Immersive)	7.2500	4.39946	20
	Stereo screen (Non-immersive)	9.6818	2.31735	22
	Total	8.5238	3.63746	42
Total	HMD (Immersive)	9.0882	4.27372	34
	Stereo screen (Non-immersive)	10.2353	2.52333	34
	Total	9.6618	3.53072	68

simulation were quite similar (women’s mean on the performance test was 7.71, SD = 1.61; men’s mean was 7.54, SD = 1.1;  $F = 0.23$ ,  $p = 0.63$ ). Nor did the level of immersion of the simulation have an effect on learning (mean test score in the immersive group was 7.67, SD = 1.32; mean test score in the non-immersive group was 7.62, SD = 1.56;  $F = 0.01$ ,  $p = 0.91$ ). Women and men did not differ with regard to the level of immersion of the system (the interaction between the variables was not significant ( $F = 0.11$ ,  $p = 0.73$ )) (Table 2).

**Table 2.** Performance of women and men in immersive and non-immersive systems

Gender	Immersion	Mean performance	SD	N
Men	HMD (Immersive)	7.5000	1.09193	14
	Stereo screen (Non-immersive)	7.5833	1.16450	12
	Total	7.5385	1.10384	26
Women	HMD (Immersive)	7.8000	1.47256	20
	Stereo screen (Non-immersive)	7.6364	1.76056	22
	Total	7.7143	1.61224	42
Total	HMD (Immersive)	7.6765	1.31933	34
	Stereo screen (Non-immersive)	7.6176	1.55728	34
	Total	7.6471	1.43272	68

The correlation between usability and performance was not significant, either in the whole sample ( $r = -0.05$ ,  $p = 0.68$ ), or when segmenting the sample by gender (men:  $r = -0.09$ ,  $p = 0.67$ ; women:  $r = -0.01$ ,  $p = 0.94$ ) indicating that performance was not influenced by usability.

## 4 Discussion

As in previous studies [6], no overall differences in usability were found between immersive and non-immersive systems; however, analysing the data separately for men

and women, women did present differences. The usability scores of both systems differed little, but women considered the usability of the non-immersive system to be higher. As in previous studies [8], women gave lower scores of usability than men on all the conditions, but the greatest difference was observed in the immersive system.

Future studies should explore the reasons for the differences in the usability of immersive and non-immersive VR systems between men and women. Possibly, one of those reasons could be a higher vulnerability in women to the adverse effects of some VR systems such as simulator sickness, a form of motion sickness induced by VR environments. The signs and symptoms of motion sickness include cold sweating, pallor, nausea and, in some cases, vomiting. Simulator sickness may also include other symptoms such as disorientation, disturbances to balance, eyestrain, blurred vision, drowsiness and lack of coordination. This is one of the most notable concerns related with the use of HMDs. In a study by Davis et al. [10], for example, also using Oculus Rift (the DK1 version, not the DK2 version), eight out of 12 participants reported mild levels of nausea, two moderate, and two high. In the same study, another condition produced even worse motion sickness symptoms: all participants reported some degree of nausea, with seven experiencing moderate levels and five high levels. Eight (66%) of the participants were unable to complete the study.

In the DK2 version (used in our study), Oculus partially corrects this problem by replacing the LCD screen with an OLED screen, thus achieving a higher refresh rate, and by incorporating positional tracking. In any case, despite technical improvements, the use of immersive devices such as HMDs is still associated with simulator sickness in a larger proportion of people than other less immersive devices. This point should be taken into consideration, especially when (as in our study) the use of these devices for training purposes requires long periods of time, which increases the likelihood of simulator sickness or other similar sources of discomfort. Negative side effects of this kind are extremely rare when the training is carried out using non-immersive devices such as laptops or desktop computers with stereoscopic display.

In our study, only two participants were unable complete the experiment; in one of the cases, a woman, this was due to simulator sickness. Thus it is possible that the lower scores of usability recorded by women, especially in the immersive condition, were related to mild forms of motion sickness. This hypothesis remains speculative because no measure of motion sickness was used in our study; nonetheless, it is a well-established fact that women are more susceptible than men to motion sickness in general and to simulator sickness in particular. Studying seasickness among more than 20,000 passengers on ferries, Lawther and Griffin [11] found that the severity of seasickness symptoms was greater in women than in men by a ratio of 5:3. A similar ratio was observed for vomiting, which was more common among women (8.8%) than among men (5.0%). Similar results have been obtained in land transportation [12–14] and in vehicle simulators [15]. Women are also more likely than men to experience motion sickness resulting from wind-induced motion in skyscrapers [16]. The differences between men and women in susceptibility to motion sickness extend to visual motion stimuli in the absence of inertial displacement [17]. Specifically regarding simulator sickness, several studies show that higher intensity symptoms are found in women [18–21].

Taken together, these data suggest that the higher proneness to simulator sickness in women may well be the cause of the lower scores of usability that they give to immersive VR environments compared with men. This hypothesis should now be tested in a replication of the present study, including measures of simulator sickness.

In any case, the results of this study reveal that gender differences must be taken into account when using VR environments to train professionals (in our case, to train health sciences students in diagnostic skills). Highly immersive devices may not always be the best choice; in some cases they may offer significant advantages, but sometimes “less is more” and lower levels of immersion can prevent the emergence of undesirable secondary effects such as simulator sickness, even though this may mean having to forgo some of the advantages of the immersive environments.

Virtual Reality simulations are engaging and facilitate comprehension by the means of situating learning materials in a context. Learning in a VR environment can be more effective and motivating than traditional classroom practices [22–24]. Some of these advantages are associated with their degree of immersiveness, but the importance of achieving a balance between these characteristics and the usability of the system must be taken into account. Individual differences, including gender, appear to be an important factor in appreciations of usability.

**Acknowledgments.** This study was supported by the Spanish Ministry of “Economía y Competitividad” (Project PSI2015-70389: ‘Desarrollo de técnicas de exposición mediante realidad virtual para la mejora del tratamiento de la anorexia nerviosa’), and by “Programa de Millora i Innovació Docent de la Universitat de Barcelona” (PMID-UB)

## References

1. Riva, G., Baños, R., Botella, C., Mantovani, F., Gaggioli, A.: Transforming experience: the potential of augmented reality and virtual reality for enhancing personal and clinical change. *Front. Psychiatry* **7**, 164 (2016). doi:[10.3389/fpsy.2016.00164](https://doi.org/10.3389/fpsy.2016.00164)
2. Ferrer-García, M., Gutiérrez-Maldonado, J., Caqueo-Urizar, A., Moreno, E.: The validity of virtual environments for eliciting emotional responses in patients with eating disorders and in controls. *Behav. Modif.* **33**(6), 830–854 (2009). doi:[10.1177/0145445509348056](https://doi.org/10.1177/0145445509348056)
3. Gutiérrez-Maldonado, J., Ferrer-García, M., Caqueo-Urizar, A., Letosa-Porta, A.: Assessment of emotional reactivity produced by exposure to virtual environments in patients with eating disorders. *CyberPsychology Behav.* **9**(5), 507–513 (2006)
4. Bell, J.T., Fogler, H.S.: The investigation and application of virtual reality as an educational tool. In: *Proceedings of the American Society for Engineering Education* (1995)
5. Gutiérrez-Maldonado, J., Ferrer-García, M., Dakanalis, A., Riva, G.: Virtual reality: applications to eating disorders. In: Agras, S. (ed.) *Oxford Handbook on Eating Disorders*, 2nd edn. (in press)
6. Gutiérrez-Maldonado, J., Ferrer-García, M., Pla-Sanjuanelo, J., Andres-Pueyo, A., Talarn-Caparrós, A.: Virtual reality to train diagnostic skills in eating disorders. Comparison of two low cost systems. *Stud. Health Technol. Inform.* **219**, 75–81 (2015)
7. Gutiérrez-Maldonado, J., Ferrer-García, M.: Are virtual patients effective to train diagnostic skills? A study with bulimia nervosa virtual patients. In: Magnenat-Thalmann, N., Wu, E., Tachi, S., Thalmann, D., Porcher Nedel, L., Xu, W. (eds.) *VRST*, p. 267. ACM (2013)



8. Gutiérrez-Maldonado, J., Andrés-Pueyo, A., Talarn-Caparrós, A., Achotegui-Loizate, J.: Virtual reality for training diagnostic skills in eating disorders. Gender and usability. In: 18th International Conference on Human-Computer Interaction, Toronto, Canada (2016)
9. Kirakowski, J., Corbett, M.: SUMI: the software measurement inventory. *Br. J. Educ. Technol.* **24**, 210–212 (1993)
10. Davis, S., Nesbitt, K., Nalivaiko, E.: Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. In: Proceedings of the 11th Australasian Conference on Interactive Entertainment, Sydney, Australia, vol. 167, pp. 3–14. CRPIT (2015)
11. Lawther, A., Griffin, M.J.: A survey of the occurrence of motion sickness amongst passengers at sea. *Aviat. Space Environ. Med.* **59**, 399–406 (1988)
12. Golding, J.F.: Motion sickness susceptibility. *Auton. Neurosci. Basic Clin.* **129**, 67–76 (2006)
13. Park, A.H.-Y., Hu, S.: Gender differences in motion sickness history and susceptibility to optokinetic rotation-induced motion sickness. *Aviat. Space Environ. Med.* **70**, 1077–1080 (1999)
14. Turner, M., Griffin, M.J.: Motion sickness in public road transport: the relative importance of motion, vision, and individual differences. *Br. J. Psychol.* **90**, 519–530 (1999)
15. Kennedy, R.S., Lanham, D.S., Massey, C.J., Drexler, J.M., Lilienthal, M.G.: Gender differences in simulator sickness incidence: implications for military virtual reality systems. *Safe J.* **25**, 69–76 (1995)
16. Lamb, S., Kwok, K.C.S., Walton, D.: Occupant comfort in windexcited tall buildings: motion sickness, compensatory behaviors, and complaint. *J. Wind Eng. Ind. Aerodyn.* **119**, 1–12 (2013)
17. Koslucher, F.C., Haaland, E., Malsch, A., Webeler, J., Stoffregen, T.A.: Sex differences in the incidence of motion sickness induced by linear visual oscillation. *Aviat. Med. Hum. Perform.* **86**, 787–793 (2015)
18. Hein, C.M.: Driving simulators: six years of hands-on experience at hughes aircraft company. In: Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, pp. 607–611 (1993)
19. Biocca, F.: Will simulation sickness slow down the diffusion of virtual environment technology? *Presence: Teleoperators Virtual Environ.* **1**, 334–343 (1992)
20. Flanagan, M.B., May, J.G., Dobie, T.G.: Sex differences in tolerance to visually-induced motion sickness. *Aviat. Space Environ. Med.* **76**(7), 642–646 (2005)
21. Park, G.D., Allen, R.W., Fiorentino, D., Rosenthal, T.J., Cook, M.L.: Simulator sickness scores according to symptom susceptibility, age, and gender for an older driver assessment study. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, pp. 2702–2706 (2006)
22. Fominykh, M., Prasolova-Førland, E., Morozov, M., Smorkalov, A., Molka-Danielsen, J.: Increasing immersiveness into a 3D virtual world: motion-tracking and natural navigation in vAcademia. *IERI Procedia* **7**, 35–41 (2014). doi:[10.1016/j.ieri.2014.08.007](https://doi.org/10.1016/j.ieri.2014.08.007)
23. Monahan, T., McArdle, G., Bertolotto, M.: Virtual reality for collaborative e-learning. *Comput. Educ.* **50**(4), 1339–1353 (2008)
24. Trindade, J., Fiolhais, C., Almeida, L.: Science learning in virtual environments: a descriptive study. *Br. J. Educ. Technol.* **33**(4), 475–492 (2002)