SOCIEDAD CIENTÍFICA ESPAÑOLA PARA LA INVESTIGACIÓN Y FORMACIÓN EN CIENCIAS DE LA SALUD

DUC IN ALTUM
UA
UNIVERSIDAD AUTÓNOMA DE CHILE

European Journal of Education and Psychology
2026, Vol. 19, Nº 1 (Págs. 1 - 21)

Eur. j. educ. psychol. e-ISSN 1989-2209
https://revistas.uautonoma.cl/index.php/ejep
doi: 10.32457/ejep.v19i1.3416

Virtual Reality Learning Objects: Experimental Comparison of Immersive and Desktop Modalities on Academic Performance and Usability

Objetos de aprendizaje en realidad virtual: comparación experimental de modalidades inmersiva y de escritorio en rendimiento y usabilidad

Julio Cabero-Almenara[1]; María Miravete-Gracia[2];
Cristóbal Ballesteros-Regaña[1]*; Antonio Palacios-Rodríguez[1]
[1]Departamento de Didáctica y Organización Educativa, Universidad de Sevilla, España;
[2]Universidad Internacional de Andalucía, España.

****Corresponding author:*** cballesteros@us.es

ABSTRACT
This study investigates virtual reality (VR) learning objects by comparing immersive and desktop modalities in terms of academic performance and perceived usability. Using an experimental pretest–posttest design, 336 undergraduate education students were randomly assigned to immersive (Meta Quest 3) or desktop VR conditions. Usability was assessed through the System Usability Scale (SUS), while learning outcomes were measured via objective tests. Results indicate significant learning gains in both modalities, with higher posttest scores favoring the desktop condition, although effect sizes were moderate. Perceived usability was rated as notable in both formats, with no statistically significant differences, despite a slight preference for immersive experiences. A positive and moderate correlation between usability and academic performance was identified. The findings confirm that both VR delivery modes are effective for knowledge acquisition, highlighting the critical role of instructional design principles and time management in optimizing learning outcomes.

Keywords: Virtual Reality; Immersive Learning; Desktop-Based Learning; System Usability Scale; Learning Outcomes; Educational Technology; Experimental Design

RESUMEN

Este estudio analiza los objetos de aprendizaje en realidad virtual (RV) mediante la comparación de modalidades inmersiva y de escritorio en términos de rendimiento académico y usabilidad percibida. A través de un diseño experimental con pretest y postest, 336 estudiantes universitarios de educación fueron asignados aleatoriamente a condiciones de RV inmersiva (Meta Quest 3) o de escritorio. La usabilidad se evaluó mediante la System Usability Scale (SUS) y el aprendizaje mediante pruebas objetivas. Los resultados evidencian mejoras significativas en ambas modalidades, con puntuaciones superiores en el grupo de escritorio, aunque con tamaños del efecto moderados. La usabilidad fue valorada como notable en ambos formatos, sin diferencias estadísticamente significativas, aunque con ligera preferencia por la experiencia inmersiva. Se identificó una correlación positiva moderada entre usabilidad y rendimiento. Los hallazgos confirman la eficacia de ambas modalidades de RV para la adquisición de conocimientos, destacando la importancia del diseño instruccional y la gestión del tiempo.

Palabras Clave: Realidad Virtual; Aprendizaje Inmersivo; Aprendizaje en Escritorio; Escala de Usabilidad del Sistema; Resultados de Aprendizaje; Tecnología Educativa; Diseño Experimental

INTRODUCTION

Augmented Reality (AR) and Virtual Reality (VR) are technologies that are currently being adopted at a rapid pace in educational institutions, particularly universities, radically transforming the way knowledge is imparted and learning is experienced through immersive environments and interactive simulations (Panerai *et al.,* 2018; Wang & Li, 2024).

These technological innovations enable the creation of immersive experiences that progressively blur the boundaries between the physical world and digital environments, offering more immersive forms of interaction that are adapted to human behavior (Bartu & Bayraktar, 2024).

Regarding its conceptualization, the proposals made by different authors are diverse (Cabero-Almenara, 2023) and present some conceptual confusion. To clarify them, Rauschnabel *et al.* (2022) present the following diagram (Figure 1).

Figure 1.

Diagram of mixed or extended reality, AR, and VR. Source: Rauschnabel et al. (2022, p.6)

If we specify the meaning of each technology, we can state that AR superimposes virtual elements onto the real world, enriching the perception of the environment without isolating the user from their physical context. As for VR, it can be understood as the sum of hardware and software systems that seek to perfect a complete sensory illusion of being immersed and submerged in a virtual and digital environment (Lai and Cheong, 2022; Cabero-Almenara, 2023).

Based on their modes of use, Lai and Cheong (2022) establish three categories: (a) non-immersive VR, which is the most common form and involves a motion sensor that detects the user's movement and reproduces it on screen to create a virtual world; b) semi-immersive VR, which provides a partially virtual environment and is commonly used for educational and training purposes (such as flight simulators); and c) fully immersive VR, which provides users with the most realistic experience possible. In this case, special head-mounted displays are required to provide sensory content with a wide field of view and can even be programmed to provide full-body haptic feedback.

For their part, various authors (Caballero *et al.,* 2020; Mulders *et al.*, 2020; Cigdem *et al.,* 2024) establish a two-fold distinction: non-immersive or desktop-based (which relies on traditional devices such as a mouse and keyboard) and high-immersion, which generally involves the use of special glasses, of which there are different types (Swidrak, 2023).

Although most research suggests that incorporating VR into teaching improves student performance (De Back *et al.,* 2020; Agurto-Cabrera and Guevara-Vizcaíno, 2023; Victoria-Maldonado *et al.,* 2024; Rzanova *et al.,* 2023); however, studies currently underway to compare both forms of use show contradictory results. Various meta-analyses conducted on the same (Hamilton *et al.,* 2021; Iglesias, 2022) demonstrate that the immersive form of use was more effective, adding that the effect sizes were small.

However, other meta-analyses conducted on the desktop version (Gonzáles *et al.*, 2024) indicated that its use can effectively improve students' academic performance, fostering improvements in their knowledge, emotions, and skills; and we also found studies demonstrating the absence of significant differences between both forms of use (Mariscal *et al.,* 2020; Yudintseva, 2024). Consequently, as Roda-Segarra *et al.* (2022) point out, experiences vary depending on the type of software used.

System Usability: The System Usability Scale (SUS)

Usability is defined as the capacity of a specific tool or object to be used efficiently and effectively. According to Torres (2018), this concept is related to several elements, including ease of use, user experience, and user-centered design.

Although the procedures for its evaluation are diverse and range from "heuristic evaluation" (with usability experts), "user testing" (to gather users' responses during interaction), "user interviews" (to collect their experiences, opinions, and feelings), performing specific actions on the object, or analyzing "usage metrics" generated by users during interaction; nevertheless, the most commonly used techniques are "usability surveys" conducted via structured questionnaires designed to collect data on user satisfaction, ease of use, and other aspects of the experience during the user's interaction with the technological object. Although a variety of surveys have been proposed for this purpose (Tobón *et al.,* 2017; Ramos, 2023), one of the most widely used is the "System Usability Scale" (SUS).

The SUS scale was designed by Brooke (1996 and 2013) and has become one of the most widely used tools for assessing a person's perceived usability of a technological object or system (Lewis, 2018; Gronier and Baudet, 2021). As evidenced by the number of references found (approximately 1,820,000) in a search conducted on April 27, 2025, on Google Scholar using the term "System Usability Scale."

Its effectiveness is demonstrated by the fact that it has been used to analyze various types of technologies, such as: medical apps (Hyzy *et al.,* 2022), mobile apps (Dege *et al.,* 2024), online training systems (Ozzi, 2024), social media (Angulo-Armenta *et al.,* 2022), audio podcasts (Mulero-Henríquez *et al.*, 2024), artificial intelligence (Artiles-Rodríguez *et al.,* 2021; Patience *et al.,* 2023), websites (Galuh *et al.,* 2021), as well as AR, VR, and extended reality objects (Fernandes *et al.,* 2021; Campo-Prieto *et al.,* 2021; Cabero-Almenara, *et al.,* 2025).

At the same time, its original version has been translated into various languages such as Spanish, Dutch (Ensink *et al.*, 2024), Portuguese (Martins *et al.,* 2017), French (Gronier and Baudet, 2021), Chinese (Wang *et al.*, 2020), Spanish (Castilla *et al.,* 2023), and Malay (Marzuki *et al.,* 2018); high levels of reliability have been obtained in various meta-analyses and literature reviews (Lewis, 2018; Vlachogianni and Tselios, 2022) which, in most cases, achieve Cronbach's alpha values above .80.

METHOD

The research design was a two-group experimental design, in which each group was administered a pretest and a posttest and received a different treatment: interaction with the desktop version and with the immersive version. In this type of design, the sample was randomly divided into two experimental groups (Arnau, 1981). The following stages were carried out to implement the study:

a) A session to explain VR to the students and their interaction with different learning objects produced by the researchers, in both the "desktop" and "immersive" versions. At the end of the session, the pretest was administered regarding the content covered in the learning object: "Classroom of the Future." The purpose of this session was to eliminate the "novelty" effect so as not to bias the results, as often occurs when using this type of technology. This session allowed for controlling the potential confounding variable of differences in prior knowledge and familiarity that students might have regarding VR technology.

b) Division of the class into two groups that would work in separate sessions, one with the desktop version of the learning object and the other with the immersive version. The assignment of subjects to the immersive or desktop group was random, since the original class was divided by the faculty into two practical subgroups, each comprising approximately 50% of the original group. It was decided that in all classes where the experiment was conducted, the first subgroup would participate in the immersive version and the second in the desktop version. At the end of the session, the post-test and the SUS usability scale were administered. The immersive version was conducted using Meta Quest 3 headsets (Figure 2).

Figure 2.
Images of the experiment

Objectives

a) To determine whether there were significant differences in the perceived usability of a learning object created in VR format, as determined by the version—desktop or immersive—in which it was used.

b) Analyze whether the version—immersive or desktop—influenced students' acquisition of information from the learning object produced in VR format.

c) Determine whether there is a significant relationship between performance and the perceived usability of the learning object produced in VR format

Participants

The research sample consisted of 336 students enrolled in the course "ICT Applied to Education" in the elementary and early childhood education programs offered by the Faculty of Education Sciences at the University of Seville. Of these, 165 (49.1%) interacted with the desktop version and 171 (50.9%) with the immersive version. 21.7% (n=73) were male and 78.3% (n=263) were female. The assignment of participants to the immersive or desktop group was random, as the original class was divided by the faculty into two practical subgroups, each comprising approximately 50% of the original group. It was decided that in all classes where the experiment was conducted, the first subgroup would participate in the immersive version and the second in the desktop version.

The study was conducted within the context of regular classroom instruction, as part of the subject's standard curriculum and the activities outlined in the course syllabus, and was not introduced as a specific intervention. The study was educational in nature, low-risk, non-interventional, and classroom-based. Furthermore, it should be noted that the questionnaire included an informed consent form for the students. Data confidentiality was maintained in accordance with the standards applied to all student data collected in class.

Instruments

Two instruments were used to collect the information. For academic performance, an "ad hoc" instrument was used in a pretest and posttest version. The pretest, in addition to collecting biographical data on the student, included 25 multiple-choice questions with a single correct answer regarding the content covered in the learning objectives. The posttest included the same questions as the pretest but with the order and answer options changed; however, to ensure that the instrument did not lead to "overlearning," the pretest was administered approximately three weeks before the study was conducted.

As we have previously noted in another study (Cabero-Almenara, *et al.,* 2025), to analyze usability, we employed the "SUS" scale due to its ease of application and comprehension by the study participants.

To assess the reliability of the SUS scale, Cronbach's alpha and McDonald's omega were used (Orcan, 2023; Colorado *et al.,* 2024). Following its application, a value of .901 was obtained for Cronbach's alpha and .898 for McDonald's omega. These values, according to Oviedo and Campo-Arias (2005), George and Mallery (2003), and Roco-Videla *et al.* (2024), suggest a high reliability index for the instrument.

Procedure: produced material-object

For the research, a learning object called "Classroom of the Future" was created. It was divided into three main parts: a) the program's title screen, a screen displaying operating instructions and the production credits; b) a screen displaying the three classrooms offered : the Faculty of Education at the University of Santiago de Compostela and the University of the Basque Country, and the Resource Center of the Teachers' Center at the University of Seville; and c) an individual presentation area for each of the aforementioned classrooms of the future (Figure 3).

Figure 3.

Program cover and screen with navigation instructions (Source: author's own work)

The research was conducted in the classroom at the University of Santiago de Compostela. For its production, the phases proposed in the ADDIE design model were followed: Analysis, Design, Development, and Evaluation (Branch, 2009), while also considering the principles proposed by Mayer's Cognitive Theory of Multimedia Learning (2004 and 2022): (a) the multimedia principle; (b) spatial and temporal contiguity; (c) modality; (d) redundancy; (e) coherence; (f) signaling; (g) segmentation; and (h) pre-training, to activate prior knowledge.

Recent research, such as that by Alpizar *et al.* (2020) and Mayer *et al.* (2023), particularly highlights the relevance of the segmentation principle, noting that it facilitates comprehension by organizing content into small, structured blocks, which is essential for capturing and maintaining student attention. Complementarily, Mulders *et al.* (2020) revisit three essential foundations for multimedia content design: (1) Dual Coding Theory, which posits that information processing occurs through two channels: verbal and visual; (2) the limited processing capacity of working memory, which necessitates avoiding cognitive overload; and (3) the need for active processing, where the learner selects, organizes, and integrates information with prior knowledge to construct meaning.

Principles of Multimedia Theory have been widely applied in recent years in the development of materials in AR and VR formats (Mulders, Buchner, & Kerres, 2020; Oje, Hunsu, & Dominik, 2023; Cabero-Almenara *et al.,* 2025).

Various software programs (Krpano, Adobe Photoshop, Adobe Premiere Pro, Insta360 Studio, and Canva) were used to produce the VR content: editing and assembling panoramic images and 360° videos, designing holograms, and incorporating interactive elements to enrich the user experience.

These theoretical guidelines translate into practical recommendations for the design of digital educational resources, such as eliminating distracting elements, using effective cues, and favoring audiovisual presentations with narration over written text, especially in materials of lower conceptual complexity. Such recommendations take on particular relevance in the context of immersive environments such as VR, where, according to Parong and Mayer (2021), the emotional impact and playful nature can divert attention from the educational objective toward a predominantly recreational experience. Therefore, it is crucial to design materials that balance interactivity and pedagogical intent, incorporating intuitive navigation resources, interactive points (hotspots) that expand information through various formats (such as still images, videos, animations, podcasts, or holographic presentations), and visual elements that guide the user's journey (Alpizar *et al.,* 2020).

A variety of technological resources were integrated into the designed virtual environment: video clips, animations, holographic presentations, multimedia, and visual sequences with progressive text. To facilitate interaction and navigation within the environment, various types of iconography were also incorporated, such as directional arrows and informational hotspots. The latter were programmed to change color upon activation, thus allowing students to easily recognize which information was being explored at any given moment (Figure 4). Specifically, the learning object incorporated twelve audio clips, four videos with text, three videos without text (two of which featured people), two holograms, and two still images with audio.

Figure 4.
Different types of resources used in the project

RESULTS

To facilitate the description and understanding of the results achieved, we will begin, first, with performance. Next, we will apply the SUS scale and conclude by analyzing whether significant differences were found between the performance achieved by the students and the usability evaluations conducted on the learning object produced, depending on the type of experience in which they participated (desktop or immersive).

Starting with performance, Table 1 presents the mean scores and standard deviations achieved, both considering the overall VR experience and differentiating the scores obtained based on the type of experience in which the students participated.

Table 1.
Academic Performance: Pre-test and Post-test. Overall Experience, and Desktop and Immersive Versions

	Overall Exp.		Desktop Exp.		Immersive Exp.	
	Mean	**Std. Dev.**	**Average**	**Est.**	**Medium**	**Est.**
Pre-test	21.60	2.621	21.87	2,217	21.33	2,940
Post-test	22.94	2,632	23.44	1,523	22.47	3,308

The results obtained show, first, that post-test scores are higher than pre-test scores, considering both the entire experience as a whole and differentiating, within it, the type of interaction carried out with the learning object. And, second, that the standard deviation scores obtained reflect a certain dispersion of the data around two points.

It should be noted that to test the various hypotheses that will be formulated, we will use nonparametric statistical tests, which have the advantage of not requiring a normal distribution and are very useful for rank-based tests, as is the case here (Politi *et al.*, 2021).

To achieve one of the objectives, we formulate the following hypotheses:

a) Null hypothesis (H0): There are no significant differences between the scores obtained on the pretest and posttest, at a significance level of $p \leq .05$.

b) Alternative hypothesis (H1): There are significant differences between the scores obtained on the pretest and posttest, at a significance level of $p \leq .05$.

To test these hypotheses, we used the nonparametric Wilcoxon statistical test. Table 2 presents the results obtained both overall and for the different interaction experiences with the object.

Table 2.

*Wilcoxon statistic (note: **= significant at p ≤ .001)*

Overall Experiment		Desktop Experience		Immersive Experience	
Wilcoxon	Sig	Wilcoxon	Sig	Wilcoxon	Sig
-9.762	.001(**)	-7.786	.001(**)	-6,072	.001(**)

The results allow us to reject the null hypothesis (H0) at a significance level of $p \leq .01$. Consequently, it can be concluded that the students' performance showed significant differences between their pretest and posttest scores.

When analyzing the effect size using Cohen's d, the values were 0.533 for the overall study, 0.473 for the immersive group, and 0.595 for the desktop group. This suggests that the effect sizes obtained in all cases can be considered intermediate or moderate (López-Martín and Ardura-Martínez, 2023) .

To determine whether the differences favored the scores obtained in the pretest or posttest, the rank sum test was applied. Table 3 presents the scores obtained.

Table 3.

Rank test between pretest and posttest scores

Experience	Rank type	Average rank	Sum of ranks
Overall	Negative ranks	110.18	5729.50
	Positive ranks	140.35	30,316.50
Desktop	Negative ranges	43.68	961.00
	Positive ranges	70.50	7,685.00
Immersive	Negative ranges	64.27	1928.00
	Positive ranges	70.33	7525.00

The values obtained in the sum of ranks clearly show that the scores achieved by the students on the posttest are higher than those obtained on the pretest.

Consequently, and taking into account the values obtained in both the Wilcoxon test and the rank sum test, it can be concluded that there are significant differences between the pretest and posttest scores, in favor of the latter, which indicates that the students acquired the content presented in the learning object produced, both overall and through the experiences carried out in the desktop and immersive versions. This suggests that the use of the produced learning object has facilitated the students' acquisition of the information.

To further explore the objectives related to academic performance, the following hypotheses were formulated:

a) Null hypothesis (H0): There are no significant differences between the use of the VR object produced in a desktop or immersive version in the pretest and posttest, at a significance level of $p \leq .05$.

b) Alternative hypothesis (H1): There are significant differences between the use of the VR object produced in a desktop or immersive version in the pretest and posttest, at a significance level of $p \leq .05$.

To this end, we applied the Mann-Whitney U test. Table 4 presents the differences between the desktop and immersive versions.

Table 4.

*Mann-Whitney U test (note: ** = significant at p ≤ .001)*

	Pre-test_score	**Post-test score**
Mann-Whitney U	13159.000	11,656.000
Wilcoxon W	27,865.000	26,362.000
Z	-1,080	-2,816
Sig. asin. (bilateral)	.280	.005 (**)

The results obtained show, on the one hand, that there were no significant differences at $p \leq .01$ between the desktop and immersive groups in the pretest scores; that is, both groups started from a similar level of knowledge regarding the content presented in the VR object produced, and therefore H0 was not rejected, which also supported one of the characteristics of this type of design, namely that the groups were equivalent (Arnau, 1981). Conversely, if such differences were found in the post-test scores at ≤ .01, then, as a consequence, in this case H0 is rejected and H1 is accepted.

When analyzing the effect size, the value obtained was $r = 0.154$, which can be considered a small effect size.

In order to analyze which type of experience (desktop or immersive) accounted for this difference, the rank test was applied, yielding the values shown in Table 5. The differences favored the desktop version.

Table 5.
Values obtained in the rank test

	Type	N	Average rank	Sum of ranks
Post-test score	Desktop	165	183.36	30,254.00
	Immersive	171	154.16	26,362.00
	Total	336		

Moving on to the objective regarding the system's usability, Table 6 presents the mean values and standard deviations obtained for both the different items of the instrument and the overall score achieved on the scale.

Table 6.
Means and standard deviations of the SUS scale items and the total instrument score

	Overall		**Desktop**		**Immersive**	
Items	**Mean**	**D. est.**	**Average**	**Est.**	**Average**	**Est.**
I think I would like to use this system frequently.	4.13	.854	4.05	.878	4.20	.825
I found the system unnecessarily complex.	2.43	1.12	2.52	1.09	2.35	1.10
I think the system is easy to use.	3.91	.90	3.82	.93	3.99	.86
I think I would need help from someone with technical knowledge to use this system.	3.33	1.19	3.31	1.19	3.35	1.19
I found that the various features of this system were well integrated.	4.16	.79	4.07	.85	4.24	.71
I think the system is very inconsistent.	2.12	1.08	2.17	1.12	2.07	1.04
I imagine most people would learn to use this system very quickly.	3.63	1.04	3.45	1.10	3.81	.97
I find the system very difficult to use.	2.24	1.14	2.25	1.11	2.22	1.17
I felt very safe using the system.	3.78	.93	3.80	.91	3.76	.94

	Overall		**Desktop**		**Immersive**	
Items	**Mean**	**D. est.**	**Average**	**Est.**	**Average**	**Est.**
I needed to learn a lot of things before getting started with the system.	2.43	1.21	2.53	1.27	2.33	1.15
SUS	67.22	11.27	66.25	11.83	68.20	10.70

It is important to note that when determining the usability level of an object or instrument, one must consider that, although the scale is structured as a standardized questionnaire consisting of 10 Likert-type items, with values ranging from 1 (strongly disagree) to 5 (strongly agree), the SUS score is not calculated directly. To obtain it, as Denzin (2017) notes, the scores of the odd-numbered items are added together, and five units are subtracted from this total; simultaneously, the scores of the even-numbered items are added together, and the resulting total is subtracted from 25. Finally, the sum of both scores is multiplied by 2.5 to obtain the final score of the scale.

The values obtained, and in accordance with the proposal made by Gimeno (2018), who considers that a score below 50 should be considered insufficient; above 80, outstanding; between 50 and 65, adequate (good); and between 65 and 80, notable. It can be noted that the students considered that the VR learning object developed, used in both the desktop and immersive versions, possessed an adequate level of usability.

In order to verify whether there were significant differences in the students' ratings of the system's usability when the VR-based object was used in the desktop and immersive versions, the following hypotheses were formulated:

a) Null hypothesis (H0): There are no significant differences in the students' ratings regarding the usability of the VR object used in the desktop or immersive versions, at a significance level of $p \leq .05$.

b) Alternative hypothesis (H1): There are significant differences in the students' ratings of the usability of the VR object used in the desktop or immersive version, at a significance level of $p \leq .05$.

To this end, the Mann-Whitney U test was applied, yielding the scores presented in Table 7.

Table 7.
Wilcoxon's W

	12640.500
Wilcoxon's W	26335.500
Z	-1.648
Significance (two-tailed)	.099

The scores obtained do not allow us to reject H0 regarding the absence of significant differences at a p-level of ≤ .05 between the two versions. Consequently, it can be noted that the usability levels found in the use of the VR object in the desktop and immersive versions were statistically similar.

Finally, we analyzed whether there was a relationship between the scores assigned by the students on the SUS usability scale and the academic performance achieved on the posttest. To do this, we applied Spearman's Rho statistical test, yielding the scores presented in Table 8.

Table 8.

Correlation coefficient between post-test scores and usability

	Spearman's Rho	Significance
Overall	.443	<.001
Desktop version	.430	<.001
Immersive version	.454	<.001

The values obtained reveal a number of facts: a) the scores achieved are positive; consequently, when one score increases, so does the other; b) the values achieved are significant; and c) a moderate correlation has been obtained, in accordance with Mateo (2004).

DISCUSSION AND CONCLUSIONS

The conclusions of the research point in different directions. The first of these concerns the high levels of reliability achieved with the instrument used to analyze the usability of the technological object produced; that is, the "System Usability Scale" (SUS). More specifically, it can be noted that the reliability index found in this study is consistent with those obtained by other authors such as Lima *et al.* (2021) and Balanyà Rebollo and De Oliveira (2024).

This finding, coupled with the fact that it is easy and quick to administer, makes it a useful and valid tool for analyzing the usability of a technological resource in general, and specifically for AR, VR, and Extended Reality (XR) resources; where, moreover, its use is gaining ground, as seen in the research by Lima *et al.* (2021), Asghar *et al.* (2023), and Cabero-Almenara *et al.* (2025).

Another finding from the study shows that VR-based learning objects are effective in helping students acquire the information presented in them, consistent with the results of previous studies such as those by De Back *et al.* (2020), Agurto-Cabrera and Guevara-Vizcaíno (2023), Victoria-Maldonado *et al.* (2024), and Rzanova *et al.* (2023). This was true regardless of whether they interacted with the object in desktop mode or in immersive mode using the "Meta Quest 3" headset. The distinction between these two modes of use has already been identified in previous research, such as that by Mariscal *et al.* (2020) and Yudintseva (2024).

In any case, in the present study, these effects were generally considered moderate, suggesting that other variables may influence these findings, such as the cognitive load incurred by the student or their intrinsic motivation. This opens up a future line of research. These results align with those reported by other authors (Hamilton *et al.*, 2021; Iglesias, 2022)

In any case, as various studies also note (Flores *et al.*, 2014; Scavarelli *et al.*, 2021; Bartu & Bayraktar, 2024), the desktop version was perceived by students as easier to use.

This lack of difference could be due to various reasons. On the one hand, the quality with which the material was designed and produced (performing well in both versions of use) and, on the other hand, although the same amount of time was allocated to the sessions organized for interaction, the experience showed that the immersive version required more time. All of this leads to the proposal to replicate the research, but taking the noted variables into account. Furthermore, these findings are consistent with other studies (George *et al.*, 2026).

As for the quality of the produced resource, this is reinforced by the fact that the score achieved on the system usability scale indicated notable usability of the resource, regardless of its mode of use: desktop or immersive. This again points to the technical and pedagogical quality with which the resource was produced.

The study also calls for VR object design to incorporate the ADDIE model and the Cognitive Theory of Multimedia Learning (Mayer, 2004 and 2022; Branch, 2009; Alpizar *et al.*, 2020). This is reinforced by the relationship between the academic performance achieved by students and the ratings given on the usability scale.

As future lines of research, it is suggested to replicate the study (taking into account the elements mentioned above) in collaboration with the University of the Basque Country, but with objects designed for other scientific disciplines, extending the training and interaction period for students with the objects, as well as relating the usability variable to other types of variables such as the degree of technology acceptance and the cognitive load students experience when interacting with them.

Finally, it should be noted that further research is needed on the possibilities and limitations offered by both modes of use (immersive and desktop), as very few comparative studies have been conducted to date.

Acknowledgments

This research article is linked to the State Program to Promote Scientific and Technological Research and its Transfer, within the framework of the State Plan for Scientific, Technical, and Innovation Research 2021–2023. Ministry of Science and Innovation. Reference number: PID2022-136430OB-I00.

REFERENCES

Agurto-Cabrera, J. C., y Guevara-Vizcaíno, C. F. (2023). Realidad virtual para la mejora del rendimiento académico en estudiantes de educación superior. *Revista Metropolitana de Ciencias Aplicadas, 6*(S2), 233-243.

Alpizar, D., Adesope, O. O., y Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research and Development, 68*, 2095-2119. https://doi.org/10.1007/s11423-020-09748-7

Angulo-Armenta, J., Sandoval-Mariscal, C., Torres-Gastelú, C., y García-López, R. (2021). Usabilidad de redes sociales con propósitos académicos en educación superior. *Formación Universitaria, 14*(6), 25-32. https://dx.doi.org/10.4067/S0718-50062021000600025

Arnau, J. (1981). *Diseños experimentales en psicología y educación*. Trillas.

Artiles-Rodríguez, J., Guerra-Santana, M., Aguiar-Perera, M. V., y Rodríguez-Pulido, J. (2021). Agente conversacional virtual: la inteligencia artificial para el aprendizaje autónomo. *Pixel-Bit. Revista de Medios y Educación, 62*, 107–144. https://doi.org/10.12795/pixelbit.86171

Asghar, I., Ullah, R., Griffiths, M., Warren, W., Dando, L. y Davies, J. (2023). A User-Centered System Usability Evaluation of a Virtual Reality Application Developed for Solar Farm Training. En *Proceedings of the 2023 9th International Conference on Computer Technology Applications (ICCTA '23)*. Association for Computing Machinery, New York, NY, USA, 157–165. https://doi.org/10.1145/3605423.3605456

Bartu, K. y Bayraktar, Ü. (2024). The Future of Virtual Reality and Education. *TOJET: The Turkish Online Journal of Educational Technology, 23*(3), 110-119.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer.

Caballero, M. P., Mejía, C., y Romero, J. (2020). Realidad aumentada vs. realidad virtual: una revisión conceptual. *Teknos. Revista Científica, 19*(2), 10-19. https://doi.org/10.25044/25392190.991

Cabero-Almenara, J. (2023). La realidad extendida en la educación. En J. Cabero-Almenara (Coord.), *Buenas prácticas de e-learning XXIII* (pp. 86-147). Ancypel.

Cabero-Almenara, J., Miravete Gracia, M., Serrano Hidalgo, M., y Núñez Domínguez, T. (2025). Evaluación de objetos de Realidad Virtual en la educación: análisis de la usabilidad y aspectos técnicos y estéticos por estudiantes. *Hachetetepé. Revista Científica de Educación y Comunicación, 30*, e1101. https://doi.org/10.25267/Hachetetepe.2025.i30.1101

Campo-Prieto, P., Cancela, J. M., Machado, I., y Rodríguez-Fuentes, G. (2021). Realidad Virtual Inmersiva en personas mayores: estudio de casos. *Retos, 39*, 101-105. https://doi.org/10.47197/retos.v0i39.78195

Castilla, D., Jaén, I., Suso-Ribera, C., García-Soriano, G., Zaragoza, I., Bretón-López, J., Mira, A., Díaz-García, A., y García-Palacios, A. (2023). Psychometric Properties of the Spanish Full and Short Forms of the System Usability Scale (SUS): Detecting the Effect of Negatively Worded Items. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2023.2209840

Colorado, J., Romero, M., Salazar, M., y Castillo, V. (2024). Análisis comparativo de los coeficientes Alfa de Cronbach, Omega de McDonald y Alfa Ordinal. *Revista Científica y Académica, 4*(4), 2739-2755. https://doi.org/10.61384/r.c.a..v4i4.836

De Back, T. T., Tinga, A. M., Nguyen, P., y Louwerse, M. M. (2020). Benefits of immersive collaborative learning in CAVE-based virtual reality. *International Journal of Educational Technology in Higher Education, 17*(1), 1-18. https://doi.org/10.1186/s41239-020-00228-9

Dege, T., Glatzel, B., Borst, V., Grän, F., Goller, S., Glatzel, C., Goebeler, M., y Schmieder, A. (2024). Patient-Centered Chronic Wound Care Mobile Apps: Systematic Identification, Analysis, and Assessment. *JMIR Mhealth Uhealth, 12*, e51592. https://www.google.com/search?q=https://doi.org/10.2196/51592

Ensink, C. J., Keijsers, N. L., y Groen, B. (2024). Translation and validation of the System Usability Scale to a Dutch version: D-SUS. *Disability and Rehabilitation, 46*(2), 395-400.

Fernandes, J., Teles, A., y Teixeira, S. (2021). An Augmented Reality-Based Mobile Application Facilitates the Learning about the Spinal Cord. *Education Sciences, 10*(12), 376. https://doi.org/10.3390/educsci10120376

Flores, J., Camarena, P. y Avalos, E. (2014). La Realidad Virtual una Tecnología Innovadora Aplicable al Proceso de Enseñanza de los Estudiantes de Ingeniería. *Apertura. Revista de Innovación Educativa, 6*(2), http://www.udgvirtual.udg.mx/apertura/index.php/apertura/article/view/547/369

Galuh, F., Fadila, G., y Adi, N. (2021). Evaluasi Usability Website Shopee Menggunakan System Usability Scale (SUS). *Journal of Applied Informatics and Computing (JAIC), 5*(2), 146-150.

George, A., Durairaj, S.K., Kandakurti, P.K., y Hazari, A. (2026). Immersive versus non-immersive virtual reality in improving upper limb function among individuals with subacute hemiplegia: a randomized controlled trial. *BMC Neurol, 26*, 60. https://doi.org/10.1186/s12883-025-04588-5

George, D., y Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Allyn y Bacon.

Gimeno, S. (2018). *Cinco formas de interpretar un SUS*. Torresburriel. https://torresburriel.com/weblog/cinco-formas-de-interpretar-un-sus

Gonzáles, H., Moreno, Y., y D'Andrea, M. (2024). Realidad virtual inmersiva como complemento en la educación odontológica: un proceso de implementación para la docencia. *Educación Médica, 25*, 100931. https://doi.org/10.1016/j.edumed.2024.100931

Gronier, G., y Baudet, A. (2021). Psychometric Evaluation of the F-SUS: Creation and Validation of the French Version of the System Usability Scale. *International Journal of Human–Computer Interaction, 37*(16), 1571-1582. https://doi.org/10.1080/10447318.2021.1898828

Hamilton, D., McKechnie, J., Edgerton, E., y Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computer Assisted Learning, 8*(1), 1–32. https://doi.org/10.1007/s40692-020-00169-2

Hyzy, M., Bond, R., Mulvenna, M., Bai, L., Dix, A., Leigh, S., y Hunt, S. (2022). System Usability Scale Benchmarking for Digital Health Apps: Meta-analysis. *JMIR Mhealth Uhealth, 10*(8), e37290. https://doi.org/10.2196/37290

Lai, J. W., y Cheong, K. H. (2022). Adoption of Virtual and Augmented Reality for Mathematics Education: A Scoping Review. *IEEE Access, 10*. https://doi.org/10.1109/ACCESS.2022.3145991

Lewis, J. (2018). The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction, 34*(7), 577-590. https://doi.org/10.1080/10447318.2018.1455307

Lima, I., Jeong, Y., Lee, Ch., Suh, G. y Hwang, W. (2021). Severity of usability problems and system usability scale (sus) scores on augmented reality (AR) user interfaces. *ICIC Express Letters ICIC International, 12*(2), 175-183. http://dx.doi.org/10.24507/icicelb.12.02.175

López–Martín, E., y Ardura–Martínez, D. (2023). El tamaño del efecto en la publicación científica. *Educación XX1*, 26(1), 9–17. https://doi.org/10.5944/educxx1.36276

Mariscal, G., Vivas-Uriasc, M. D., Redondo-Duarted, S., y Moreno-Pérez, S. (2020). Virtual Reality Simulation-Based Learning. *Education in the Knowledge Society, 21*, Article 11, 1-11. https://doi.org/10.14201/eks.20809

Martins, A. I., Rosa, A. F., Queirós, A., Silva, A., y Rocha, N. (2017). European Portuguese validation of the system usability scale (SUS). *Procedia Computer Science, 67*, 293-300.

Marzuki, M., Yaacob, N., y Yaacob, N. (2018). Translation, Cross-Cultural Adaptation, and Validation of the Malay Version of the System Usability Scale Questionnaire for the Assessment of Mobile Apps. *JMIR Human Factors, 5*(2), e10308. https://doi.org/10.2196/10308

Mateo, J. (2004). La investigación ex post-facto. En R. Bisquerra (Coord.), *Metodología de la investigación educativa* (pp. 195-230). La Muralla.

Mayer, R. E. (2004). Should there be a three strikes rule against pure discovery learning? *The American Psychologist, 59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14

Mayer, R. E. (2022). Cognitive theory of multimedia learning. En R. E. Mayer y L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press.

Mulders, M., Buchner, J., y Kerres, M. (2020). A Framework for the Use of Immersive Virtual Reality in Learning Environments. *International Journal of Emerging Technologies in Learning (iJET), 15*(24), 208–224. https://doi.org/10.3991/ijet.v15i24.16615

Mulero-Henríquez, I., Álamo-Bolaños, A., y Cobos, M. (2024). Una experiencia innovadora a través del podcast en la educación superior. *Formación universitaria, 17*(1), 23-32. https://dx.doi.org/10.4067/S0718-50062024000100023

Oje, A., Hunsu, N., y Dominik, D. (2023). Virtual reality assisted engineering education: A multimedia learning perspective. *Computers y Education: X Reality, 3*, 100033. https://doi.org/10.1016/j.cexr.2023.100033

Orcan, F. (2023). Comparison of Cronbach's Alpha and McDonald's Omega for Ordinal Data: Are They Different? *International Journal of Assessment Tools in Education, 10*(4), 709-722.

Oviedo, H., y Campo-Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría, 34*(4), 572-580.

Ozzi, S. (2024). Statistical Analysis of System Usability Scale (SUS) Evaluations in Online Learning Platform. *Journal of Information Systems and Informatics, 6*(2), 2656-4882. https://doi.org/10.51519/journalisi.v6i2.750

Panerai, S., Catania, V., Rundo, F., y Ferri, R. (2018). Remote Home-Based Virtual Training of Functional Living Skills for Adolescents and Young Adults with Intellectual Disability: Feasibility and Preliminary Results. *Frontiers in Psychology, 9*, 1-6. https://doi.org/10.3389/fpsyg.2018.01730

Parong, J., y Mayer, R. E. (2021). Learning about history in immersive virtual reality: does immersion facilitate learning? *Educational Technology Research and Development, 69*, 1433–1451. https://doi.org/10.1007/s11423-021-09999-y

Politi MT, Ferreira JC., y Patino CM. (2021). Nonparametric statistical tests: friend or foe? *J Bras Pneumol, 47*(4):e20210292. https://doi.org/10.36416/1806-3756/e20210292

Ramos, M. J. X. (2023). Diseño, validación y usabilidad de un aplicativo móvil para la enseñanza de electrocardiografía: Design, validation and usability of a mobile application for teaching electrocardiography. *Pixel-Bit. Revista de Medios y Educación, 66*, 59–85. https://doi.org/10.12795/pixelbit.95440

Rauschnabel, Ph., Felix, R., Hinsch, Ch., Shahab, H., y Alt, F. (2022). What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior, 133*, 107289. https://doi.org/10.1016/j.chb.2022.107289

Roco-Videla, Á., Aguilera-Eguía, R., y Olguin-Barraza, M. (2024). Ventajas del uso del coeficiente de omega de McDonald frente al alfa de Cronbach. *Nutrición Hospitalaria, 41*(1), 262-263. https://dx.doi.org/10.20960/nh.04879

Roda-Segarra, J., Mengual-Andrés, S., y Martínez-Roig, R. (2022). El uso de la Realidad Virtual Inmersiva en las aulas: un meta-análisis. *Research in Education and Learning Innovation Archives, 29*, 1-13. https://doi.org/10.7203/realia.29.21488

Rzanova, S., Yushchik, E., Markova, S., y Sergeeva, A. (2023). Impact of virtual reality technologies in the context of the case method on engineering students' competencies. *Education and Information Technologies, 7*, 1–19. https://www.google.com/search?q=https://doi.org/10.56028/aetr.7.1.7.2023

Scavarelli, A., Arya, A., y Teather, R. (2021). Virtual reality and augmented reality in social learning spaces: a literature review. *Virtual Reality, 25*, 257–277. https://doi.org/10.1007/s10055-020-00444-8

Swidrak, J. (2023). Realidad virtual: ayer, hoy y mañana. En M. Ruiz y T. Johnston (Eds.), *Realidad virtual y entornos inmersivos en educación superior* (pp. 19-23). Octaedro.

Tobón, M., Zarta, R., Zartha, J., Estrada, J., Díaz, J., y Gómez, U. (2017). Vigilancia tecnológica y análisis del ciclo de vida de la tecnología: técnicas de evaluación de la usabilidad, métricas y herramientas en el sector TIC. *Espacios, 38*(22), 1-28.

Torres, D. (2018). *Usabilidad. Deja de sufrir*. Anaya Multimedia.

Victoria-Maldonado, J. J., Fuentes-Cabrera, A., Fernández-Cerero, J., y Sadio-Ramos, F. J. (2024). Influencia de la Realidad Virtual en el rendimiento académico en Educación Secundaria a través de un meta-análisis. *Pixel-Bit. Revista de Medios y Educación, 71*, 107–121. https://doi.org/10.12795/pixelbit.104279

Vlachogianni, P., y Tselios, N. (2022). Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education, 54*(3), 392-409. https://doi.org/10.1080/15391523.2020.1867938

Wang, Q., y Li, Y. (2024). How virtual reality, augmented reality and mixed reality facilitate teacher education: A systematic review. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12949

Wang, Y., Lei, T., y Liu, X. (2020). Chinese System Usability Scale: Translation, Revision, Psychological Measurement. *International Journal of Human–Computer Interaction, 36*(10), 953-963. https://doi.org/10.1080/10447318.2019.1700644

Received: 2025-11-18
Accepted: 2026-05-03