Effectiveness of virtual reality in elementary school: A meta-analysis of controlled studies

Carlos A. Lara-Alvarez 1

Ezra Federico Parra-González 1

Miguel A. Ortiz-Esparza 1

Héctor Cardona-Reyes 1*

1 Center for Research in Mathematics, Zacatecas, MEXICO
* Corresponding author: hector.cardona@cimat.mx


INTRODUCTION

Technologies such as augmented reality (AR) and virtual reality (VR) have had a boom because they are increasingly accessible to users and are found in a variety of platforms ranging from smartphones to sophisticated viewers for a more immersive experience. VR is a technology that enables users to experience computer-generated environments as if they were real. It is a simulated environment created using a computer, which can be experienced through a headset or other devices that provide a fully immersive experience.

Empirical evidence shows that VR and AR improve students’ learning outcomes and enjoyment in a variety of situations. But these technologies present some subtle differences. For instance, students that use VR allocate more attention, perceive higher feelings of presence, and show more enjoyment (Huang et al., 2021). In contrast, studies show that AR is more realistic for training (Botden et al., 2007). Students that use VR present greater retention of visual information but lower retention of auditory information in comparison with students who use AR (Huang et al., 2019).
Because of the differences, choosing one of these technologies is an important decision that mainly depends on the education area. This study focuses on VR because it can represent abstract concepts and can show the world from different perspectives and scales (Curcio et al., 2016). Hence, it can be used in several education areas from science and math to social sciences and art.

Interactive VR products can be classified as tethered VR and mobile VR. A head-mounted display (HMD) is a device worn on the head that provides visual and auditory displays, typically using LCD or CRT technology to display stereo images and may include a built-in head tracker and stereo headphones. Tethered HMDs, such as the Oculus Rift DK2, offer high-quality visual and frame rate rendering for rich graphical scenes. Mobile-rendered HMDs, such as the Samsung Galaxy Gear VR and Google Cardboard, rely on smartphone sensors to enable users to explore different virtual worlds and provide wireless freedom of movement. The Google Cardboard is the simplest and most accessible HMD. Sun et al. (2022) also consider all-in-one HMDs, but for simplicity, we only focus on tethered or mobile VR based on the experimental setup.

VR systems with three degrees of freedom (3DoF) allow the user to move in three directions: up-down, left-right, and forward-backward. These systems typically include a controller with buttons and a gyroscope that detects the user's head movement. Although they provide an immersive experience, users have limited freedom of movement, meaning they cannot walk or move around the virtual space.

On the other hand, VR systems with six degrees of freedom (6DoF) allow the user to have a higher degree of freedom of movement, as they add the ability to move in three additional axes: forward, backward, and sideways. These systems typically use controllers with joysticks that allow movement and rotation in the virtual space.

The aim of this study is to draw statistical inferences regarding the effectiveness of VR in elementary school learning. By synthesizing results across studies, we seek to obtain a comprehensive understanding of the effect and to identify underlying sources of variation in the outcomes. Unlike the study conducted by Villena-Taranilla et al. (2022a, 2022b), which did not specifically target elementary school education and did not account for variations in technology (such as tethered or mobile VR applications) or the type of materials used in the control group, our study takes these factors into careful consideration. Furthermore, our study thoroughly examines primary studies based on several quality criteria to ensure a comprehensive review.

For this aim, we focus on those studies that use the pre-test-post-test control (PPC) group design, also called the pretest-posttest randomized experimental design. In this type of experiment, students are randomly assigned to either VR group (the experimental group) or a traditional learning group (the control group). The knowledge or skill is measured at least two times, once before the intervention and once after it.

**METHODS**

The method section of this paper adheres to a comprehensive framework consisting of three distinct steps:

(a) formulation of research questions,
(b) review process, and
(c) statistical methods.

This methodology aligns with the guidelines set forth in the preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 statement (Page et al., 2021), ensuring a rigorous and transparent approach to conducting the study and deriving meaningful findings. The following sections describe each of these steps.

**Formulation of Research Questions**

Research questions (RQ) are the primary motivation for the systematic review. We structure a general RQ following the universal PICO scheme (Nishikawa-Pacher, 2022). According to this scheme, we must define problem (P), intervention (I), control (C), and outcomes (O). Following this structure, the first research question is, as follows:
1. **RQ1.** Is VR (I) a more effective learning tool for elementary school students (P) compared to traditional methods (C), as measured by exam scores at the end of an instructional program (O)?

We aim to investigate the differences between tethered VR and mobile VR interventions can affect learning. There is a lack of trials that directly compare these two types of VR interventions (tethered VR or mobile VR) or the media materials (printed, multimedia, physical activities) used in the control group. Therefore, we formulate our second research question as follows,

2. **RQ2.** Does the effectiveness (O) of VR-based learning (P) vary by modifying:
   a. the DoF; i.e., 3DoF or 6DoF of VR system (I) or
   b. the media materials used for control group (C)?

### Review Process

**Delineation of the search strategy**

This step describes the planning action to answer the RQ. Once the RQs are defined, the search strategy, composed of two steps, performs a systematic search of journal articles. In the first step, the objective is to approximate the literature addressing VR as a tool for learning. To achieve this objective, we performed a query of relevant terms over five electronic databases using their search engines.

As shown in Table 1, the databases considered in this study were, as follows:

1. Web of Science,
2. Scopus,
3. IEEE Xplore,
4. ACM Digital Library, and
5. ScienceDirect.

The relevant search terms came from the key terms used in the topic area and the review's objective. Several pilot searches were necessary to refine the keywords in the search string using trial and error.

The terms whose inclusion did not yield additional articles in the automatic search were removed. The search string that identified the first set of relevant items was used:

“Virtual reality” AND (“primary school” OR “elementary school” OR “primary education”).

The purpose of including the term virtual reality in the search string is because it is considered a hardware technology approach, as opposed to terms such as educational virtual environment (EVE) or virtual learning environment (VLE), which refers to a virtual environment that is based on a certain pedagogical model and involves various didactic objectives and provides users with experiences that they would not otherwise be able to have in the physical world and results in specific learning outcomes (Mikropoulos & Natsis, 2011).

The authors conducted an initial review of the metadata and abstracts of the articles to obtain the first set of results. For the subsequent review, a subset of articles was selected based on the following inclusion criteria:

1. The article evaluates the effectiveness of VR environments.
2. The article compares VR approaches with other ones using PPC group design.
3. The population studied should be limited to elementary school students.
4. The study must have used a headset or similar hardware to provide a fully immersive experience.

In the second step, we reviewed the selected papers and excluded:

<table>
<thead>
<tr>
<th>Database</th>
<th>Parameters</th>
<th>Number of results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web of Science</td>
<td>Topic (title/abstract/keyword)</td>
<td>95</td>
</tr>
<tr>
<td>Scopus</td>
<td>Title/abstract/keyword</td>
<td>346</td>
</tr>
<tr>
<td>IEEE</td>
<td>All</td>
<td>73</td>
</tr>
<tr>
<td>ACM</td>
<td>Title/abstract</td>
<td>6</td>
</tr>
<tr>
<td>Science Direct</td>
<td>Title/abstract/keyword</td>
<td>20</td>
</tr>
</tbody>
</table>
E1. Articles not written in English.

E2. Papers with methodological inconsistencies such as insufficient sample size, lack of control groups, poorly defined research questions, inadequate statistical analyses, unreliable or invalid measures, and flawed study design.

By applying these inclusion and exclusion criteria, we aimed to ensure the quality and relevance of the studies included in our review, and to provide a comprehensive and trustworthy overview of the effectiveness of VR environments for elementary school students.

**Research execution**

To ensure a transparent and complete review reporting, Figure 1 describes the research execution by pointing out the number of articles found in each step of the search process.

**Quality criteria**

We only considered studies that verify the equivalence of groups (experimental and control) before the case study intervention. Certain issues were identified in the studies, such as conflicting information regarding the composition of the experimental and control groups (Hui et al., 2022), or significant differences in the topic or duration of exposure between the experimental and control groups (Akman & Cakir, 2020).

**Statistical Methods**

**Data synthesis**

Analyses were performed using JASP software (JASP Team, 2022). The fixed effects model was used because there are a limited number of studies available for analysis. The fixed effects model assumes that the true effect size is the same in both studies and any differences observed between the studies are due to sampling error. Forest plots were created to display the meta-analysis findings. To explore publication bias, funnel plots were created and the Egger’s test of the intercept for funnel plot asymmetry was performed.
Finally, as specific assumptions may significantly influence the results of the systematic review, we conducted a sensitivity analysis based on the variation in inclusion criteria.

Given the limited number of studies available, our objective is to investigate the impact of individual studies in RE model, specifically by identifying influential points. These influential points are characterized by their ability to significantly alter the estimated model when present or absent.

**Calculating effect sizes**

Meta-analysis is a statistical method that enables the simultaneous and quantitative evaluation of multiple studies. The dependent variable is known as effect size. Effect sizes can be used to determine the sample size for follow-up studies or examining effects across studies. There are several methods to analyze the outcomes of a PPC experiment, we use a simple analysis of final values. Let \( x_1, s_1, \) and \( n_1 \) denote the mean (M), standard deviation (SD), and the number of participants for the experimental group after the treatment. Analogously, let \( x_2, s_2, \) and \( n_2 \) be M, SD, and number of participants for the control group after the treatment.

We calculate the standardized mean difference between the two groups as,

\[
g = \frac{x_1 - x_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} 
\]

where the numerator, \( x_1 - x_2 \), is the difference between means of observations of the experimental and control groups. The denominator is the pooled SD, and \( n = n_1 + n_2 \) is the total of participants. The value (1) is known as the Hedge’s \( g \) and it can be used to compare effects across studies, even when the dependent variables are measured in different manners (Lakens, 2013). A common correction of (1) for small samples is:

\[
w = 1 - \frac{3}{4n-9} 
\]

Effect sizes are generally defined (Cohen, 2013) as small (g=0.2), medium (g=0.5), and large (g=0.8).

**RESULTS**

As shown in Figure 1, the search produced 540 results: 95 in Web of Science, 346 in Scopus, 73 in IEEE Xplore, six in the ACM Digital Library, and 20 in Science Direct. After discarding duplicates, a total of 447 articles remained. In the first selection step, 30 articles report on VR in basic education.

After a complete review of these articles, a total of six studies, including 627 participants, were selected for the final analysis. The characteristics of the participants, intervention details, and outcome measures are presented in Table 2.

**Contrasting Virtual Reality with Conventional Methods**

As a large amount of heterogeneity was found (I²=64.2%) we used random-effects modelling (restricted ML). The forest plot displayed in Figure 2 provides an overview of the effect size and corresponding confidence interval for each study.

We found that, in general VR was more effective than the control conditions in improving knowledge (standard mean difference [SMD]=0.64, 95% CI [0.47, 0.79], p<.001). Hence, the effect is medium (Cohen, 2013). Among the included studies shown in Figure 2, the first four employed mobile VR technology, while the last two utilized tethered VR. Figure 2 presents a summary of the effects categorized by technology.

---

**Table 2. Summary of the six included studies**

<table>
<thead>
<tr>
<th>Study</th>
<th>Area</th>
<th>NEG</th>
<th>NCG</th>
<th>n</th>
<th>Control</th>
<th>Hedges G</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang et al. (2019)</td>
<td>Science</td>
<td>38</td>
<td>41</td>
<td>79</td>
<td>Video</td>
<td>0.2621</td>
<td>0.2239</td>
</tr>
<tr>
<td>Demitriadou et al. (2020)</td>
<td>Math</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>Printed material</td>
<td>1.0333</td>
<td>0.4560</td>
</tr>
<tr>
<td>Degli-Innocenti et al. (2019)</td>
<td>Music</td>
<td>18</td>
<td>18</td>
<td>36</td>
<td>Loudspeaker group listening</td>
<td>1.4883</td>
<td>0.3683</td>
</tr>
<tr>
<td>Villena-Taranilla et al. (2022a)</td>
<td>History</td>
<td>53</td>
<td>45</td>
<td>98</td>
<td>Printed material</td>
<td>0.5561</td>
<td>0.2049</td>
</tr>
<tr>
<td>Amprasi et al. (2022)</td>
<td>Selective attention</td>
<td>16</td>
<td>16</td>
<td>32</td>
<td>Typical training (gym)</td>
<td>0.3025</td>
<td>0.3466</td>
</tr>
<tr>
<td>Liu et al. (2022)</td>
<td>Science</td>
<td>170</td>
<td>192</td>
<td>362</td>
<td>Classroom</td>
<td>0.6722</td>
<td>0.1080</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>305</td>
<td>322</td>
<td>627</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. EG: Experimental group & CG: Control group
Publication bias analysis

The funnel plot, shown in Figure 3, is a tool that aids in the detection of publication bias. In the absence of bias, the plot is symmetrical. Egger’s test is commonly used to assess potential publication bias. According to this test, there was no publication bias in our meta-analysis (z=-0.141; p=0.888). Upon conducting a case wise study, we have identified an influential study, Liu et al. (2022).

Comparing Results Across Different Virtual Reality Systems and Materials Used

The moderators for each study are listed in Table 3. We started by testing the effect of each moderator separately. As shown in Table 4, test results could not reject the null hypothesis, this means that removing the variables DoF and CG media materials from the model will not considerably damage the fit of the model.

DISCUSSION

The goal of the educational process is the improvement of student learning outcomes (Fernandez, 2017). Concerning the first research question, the results of this study show that students in VR learning condition obtained higher learning outcomes than students when using other approaches. There was a medium effect
Table 3. Moderators used in the analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Media materials employed for control group</th>
<th>DoF of HeadSet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang et al. (2019)</td>
<td>Multimedia</td>
<td>3</td>
</tr>
<tr>
<td>Demitriadou et al. (2020)</td>
<td>Printed material</td>
<td>3</td>
</tr>
<tr>
<td>Degli-Innocenti et al. (2019)</td>
<td>Multimedia</td>
<td>3</td>
</tr>
<tr>
<td>Villena-Taranilla et al. (2022a)</td>
<td>Printed material</td>
<td>6</td>
</tr>
<tr>
<td>Amprasi et al. (2022)</td>
<td>Equipment for physical activities</td>
<td>6</td>
</tr>
<tr>
<td>Liu et al. (2022)</td>
<td>Multimedia</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4. Moderator coefficients were examined, but none of them yielded significant results

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard error</th>
<th>z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.780</td>
<td>1.056</td>
<td>0.738</td>
<td>0.460</td>
</tr>
<tr>
<td>DoF (6)</td>
<td>-0.477</td>
<td>0.864</td>
<td>-0.552</td>
<td>0.581</td>
</tr>
<tr>
<td>Medium (multimedia)</td>
<td>-0.031</td>
<td>1.103</td>
<td>-0.028</td>
<td>0.978</td>
</tr>
<tr>
<td>Medium (printed)</td>
<td>0.254</td>
<td>0.812</td>
<td>0.312</td>
<td>0.755</td>
</tr>
</tbody>
</table>

size (M=0.641, SD=0.143) for the post-test in VR learning. These findings are aligned with studies in other different educational levels that found a medium effect size (Luo et al., 2021; Merchant et al., 2014). VR learning offers a multi-sensory, and immersive experience. That is, VR can include various hardware and software materials—e.g., haptic devices, objects, voice, and music (Prabakar et al., 2021). Adaptability, immersive, and change of scale sense are key factors that improve the effectiveness of VR over traditional learning. Kliziene et al. (2021) suggest that virtual environments offer a wide spectrum of measures and tools to maintain and enrich traditional teaching/learning according to the student’s needs and abilities. Sanchez-Vives and Slater (2005) state that VR is immersive enough to break the connection with physical space and simulate virtual presence. This immersion is often accomplished with the environment control that makes VR compatible with several pedagogical theories—e.g., constructivist, problem-based learning, cognitive development theory, and connectivism (Nisha, 2019).

Actual mobile VR is typically 3DoF while tethered VR is 6DoF. By having greater freedom of movement, users can walk and move around the virtual space in a more natural and realistic way, providing a more immersive and convincing VR experience. Regarding the second research question, which examines the impact of modifying DoF of VR system or using different media materials for the control group on the effectiveness of VR-based learning, our observations indicate that the selected studies did not reveal any statistically significant differences in either the technology of the headset or the media materials used for the control group.

Furthermore, we argue that mobile VR (3DoF) has several appealing advantages for users. It is based on mobile devices like cardboard or plastic headsets that are compatible with a wide range of smartphones. Unlike PC or console-based VR systems (6DoF) that require complex and costly setups, mobile VR allows users to immerse themselves in VR experiences anywhere, anytime. One key advantage is its accessibility. People can experience VR without investing in specialized hardware because smartphones are widely available. Mobile devices also have app stores, providing a distribution channel for VR applications and expanding the accessibility of content. Developers can create exclusive mobile VR content, offering more options and experiences for users. Mobile VR is easy to use since it does not require complicated installations. Users only need to connect their smartphone to VR headset to seamlessly enjoy virtual worlds, enhancing the user experience. Additionally, mobile VR is often more affordable than high-end systems, making it accessible to a wider audience. In summary, mobile VR offers advantages in terms of portability, accessibility, usability, content variety, and cost-effectiveness, making it an appealing choice for users seeking immersive VR experiences.

A strong agreement exists regarding the beneficial impact of VR technologies on learning outcomes in elementary school. However, the contrary findings are put forth by Abich et al. (2021) in their secondary study. It is noteworthy that Abich et al. (2021) include studies that fail to meet our inclusion criteria. For instance, these studies involve the use of collaborative VR (Hwang & Hu, 2013), evaluate motivation (Patera et al., 2008), or involve participants from a different age range (Gwee, 2013).
We observe the presence of an influential study by Liu et al. (2022). This study holds considerable influence on the effects of VR-based learning due to its inclusion of the largest number of participants (362). Consequently, the findings from this study have a substantial impact on the estimated model. We want to point out that, it is important to promote and support future research endeavors that evaluate the impact of VR on learning outcomes by comparing it with control groups using established conventional methods and practices. It is important to emphasize the need for additional studies that compare VR-based learning with conventional methods currently employed. Furthermore, conducting more comparative assessments would provide a more comprehensive and accurate comparison between these approaches.

**CONCLUSIONS**

The conducted meta-analysis combines the outcomes of six independent experimental studies, revealing that students who engage in virtual environment learning achieve higher learning scores compared to those attending conventional classes (SMD=0.64, 95% CI: [0.36, 0.92], p<001). Remarkably, DoF of the headset do not affect the effectiveness of VR-based learning. Additionally, as mobile VR systems are more widely available than high-end systems, thereby increasing accessibility to a broader audience, these affordable systems hold great potential for educational purposes and conducting further studies. Furthermore, no significant differences in scores were observed when different media materials were employed for the control group across various experimental settings.

We recommend exploring the long-term effects of VR-based learning. This investigation would delve into the lasting impact of VR-based educational experiences on students' knowledge retention, skill development, and academic performance in comparison to traditional learning methods. Furthermore, studying individual differences in VR learning would be intriguing, examining how cognitive abilities, learning styles, and spatial skills influence learning outcomes within VR environments. Moreover, it would be interesting to identify strategies to personalize VR experiences based on individual learner characteristics, enhancing the effectiveness of virtual learning environments.

**Author contributions:** All authors were involved in concept, design, collection of data, interpretation, writing, and critically revising the article. All authors approve final version of the article.

**Funding:** The authors received no financial support for the research and/or authorship of this article.

**Ethics declaration:** Authors declared that no ethics committee approval was required since the study is a review of existing literature.

**Declaration of interest:** Authors declare no competing interest.

**Data availability:** Data generated or analyzed during this study are available from the authors on request.

**REFERENCES**


☆☆☆