International Applied Linguistics Research Conference

https://doi.org/10.29140/9781763711662-6

The Impact of Virtual Reality on L2 Learning Performance: A Meta-Analysis of Effectiveness Studies from 2015–2025

Natasha Devroede[1], Dylan Dachet[2], Fabiola Camandona[3] and Ferran Suñer Munoz[4]
[1]UCLouvain
[2]UMons
[3]Università di Torino
[4]UCLouvain

Abstract

Virtual reality (VR) has become a prominent tool in language education, enabling learners to engage with the target language in immersive and interactive contexts. Despite growing research on VR in language learning, few meta-analyses exist. This study examines its impact, with immersion level, language skill, task type, and contact with non-human characters as some of the potential moderators.

Following PRISMA methodology, peer-reviewed studies and conference proceedings published between 2015 and 2025 were identified via Scopus and Web of Science. Preliminary findings suggest overall positive trends, though further analysis is needed to determine the potential of VR. This research contributes to understanding how technological affordances interact with learner characteristics and pedagogical approaches.

Keywords: foreign language acquisition, virtual reality, L2 learning performance, meta-analysis

E-mail: natasha.devroede@uclouvain.be, dylan.dachet@umons.ac.be, fabiola.camandona@unito.it & ferran.sunermunoz@uclouvain.be

In Remy Decorte, Serge Bibauw & Glenn Stockwell (Eds.)
Proceedings of the Inaugural International Applied Linguistics Conference, 2025 (pp. 33–40). Castledown.

Introduction

In recent years, VR has gained increasing attention in language learning and teaching due to its potential to provide immersive and interactive learning experiences. However, their actual effectiveness in improving learning outcomes remains a topic of ongoing debate (Dhimolea et al., 2022).

While several systematic literature reviews have synthesized research on this topic (e.g., Dhimolea et al., 2022; Haoming & Wei, 2024; Parmaxi, 2023), few meta-analyses have been conducted to date. As a result, there is still limited empirical evidence regarding the effects of VR on second language (L2) learning outcomes. This proceeding presents a preliminary meta-analysis of VR-based language learning research focusing on the following research questions:

1. What is the impact of VR-based learning environments on the learning performance of L2 learners?
2. To what extent does the level of immersion in virtual reality (high vs. low) affect L2 learning outcomes?

Theoretical Framework

There is no universally agreed-upon definition of VR in language education (Kaplan-Rakowski & Gruber, 2019). Broadly, VR refers to computer-generated 3D environments that simulate real-world interaction (Huang et al., 2020, p. 1389), as well as 360° photos or videos, such as those created with ThingLink (Cowie & Alizadeh 2022, p. 51). A key distinction is made between low-immersion VR (LiVR), typically desktop-based, and high-immersion VR (HiVR), which uses head-mounted displays for greater sensory realism (Kaplan-Rakowski & Gruber, 2019, p. 1).

Several recent meta-analyses have explored VR's impact on language learning. B. Chen et al. (2022) and J. Chen et al. (2022) reported moderate-to-strong effects on vocabulary, speaking skills, and motivation. Yu and Xu (2022) found that immersive VR tends to outperform less immersive alternatives due to increased learner engagement. Furthermore, some authors have proposed theoretically grounded and empirically proven models on immersive learning (Makransky & Petersen 2021), which aim at explaining both cognitive and affective factors at play.

Methodology

Study Identification and Selection

Two major academic databases, Scopus and Web of Science, were used to identify relevant studies. The study identification and selection process followed the PRISMA 2020 guidelines (Page et al., 2021) to guarantee transparency and replicability. Studies were screened and selected using a set of predefined inclusion criteria presented in Figure 1.

The database search yielded a total of 1,315 records. After removing 150 duplicates, 1,165 studies remained for title and abstract screening. During this phase, inter-rater reliability was established with a Cohen's kappa of 0.72, which refers to a substantial agreement (McHugh, 2012). 1,106 studies were excluded. The remaining 59 studies underwent full-text review, during which inter-rater reliability increased to Cohen's kappa = 0.83, which represents an excellent agreement (McHugh, 2012). This process initially led to a corpus of 23 studies. However, nine studies lacked a clear control group, leaving a final corpus of 14 studies. A preliminary PRISMA 2020 flow diagram is provided below (see Figure 2).

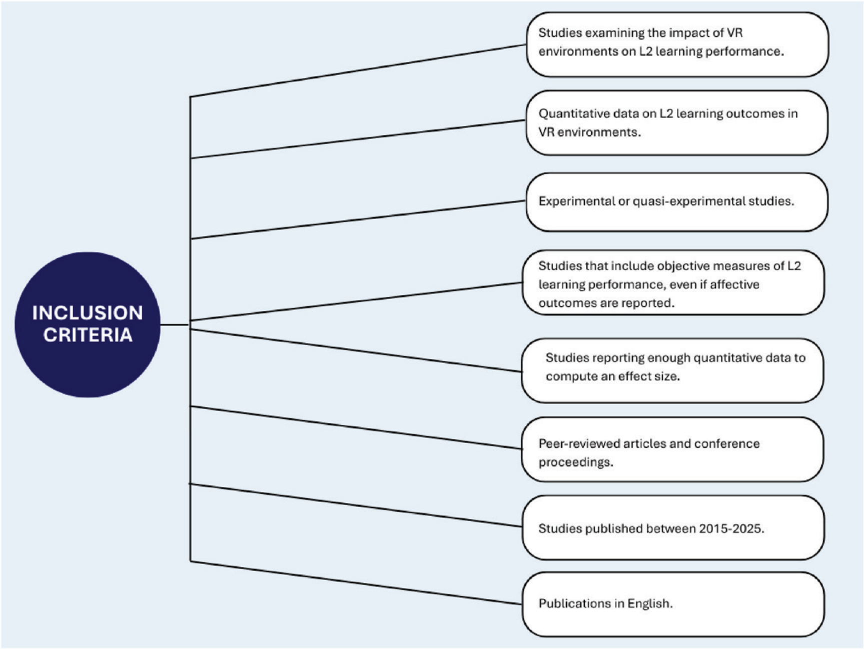

Figure 1. Inclusion criteria

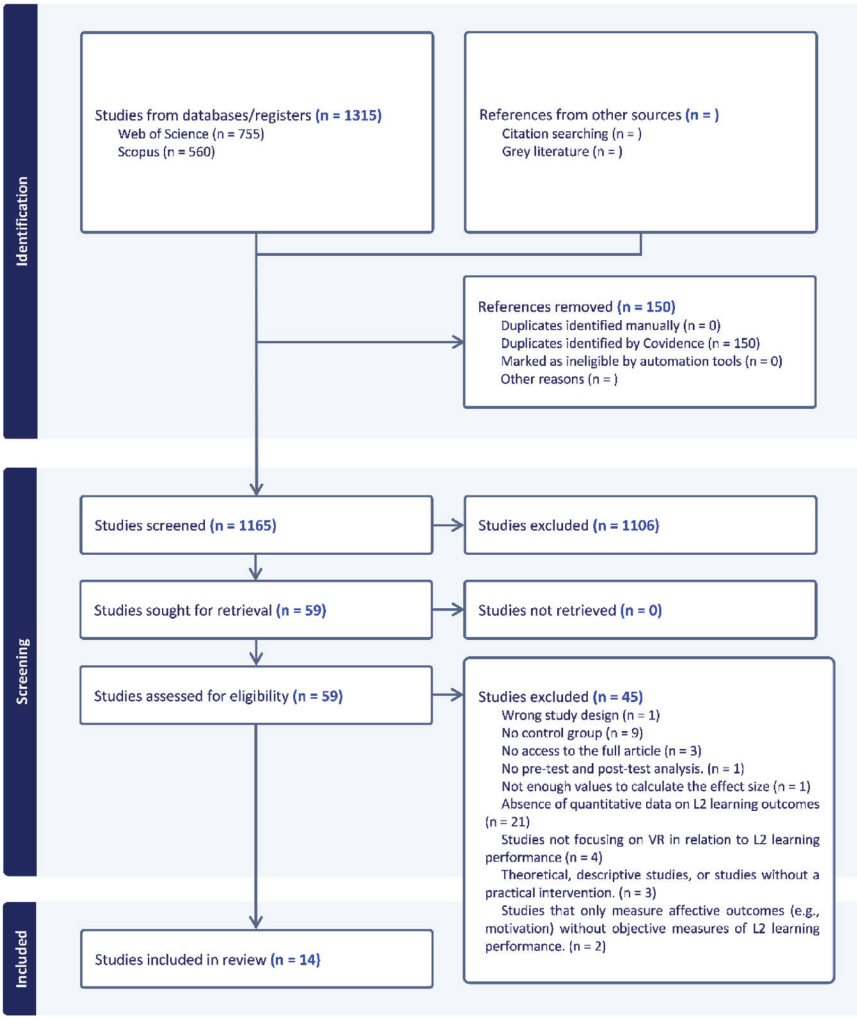

Figure 2. PRISMA Flow diagram of the article screening

Effect Size Computation and Statistical Approach

The effect sizes in this meta-analysis will be estimated using the formula of Morris (2008), with the variance of effect size computed with Lipsey and Wilson's (2001) formula and adjusted for clustering (Hedges, 2007). Analyses will be conducted in R (R Core Team, 2020) using the metafor (Viechtbauer, 2010) and clubSandwich (Pustejovsky, 2020) packages.

A Correlated Hierarchical Effects (CHE) model with Robust Variance Estimation (RVE) will be applied (Hedges et al., 2010; Pustejovsky & Tipton,

2022) to factor in dependencies between effect sizes. Meta-regression models (Schwarzer et al., 2015) will be used to examine the factors contributing to the heterogeneity of effect sizes. Among the factors studied, some relate to the quality of the studies and others to the specific characteristics of the intervention. Among the features of the study design, sample size and the type of measurement instrument is of particular importance, as it is known to have a significant effect on effect sizes (Cheung & Slavin 2012; Slavin & Madden, 2011; Slavin & Smith, 2009).

To assess publication bias, we will successively produce a funnel plot, perform an Egger test (Egger et al.,1997) and use the Trim-And-Fill (Borenstein et al., 2009) methodology. To determine the robustness of the results (Higgins et al., 2019), a leave-one-out meta-analysis will be carried out. This sensitivity analysis method consists in successively conducting multiple meta-analyses, excluding one study in each analysis in turn. A second sensitivity analysis will compare the mean effect sizes of the studies most aligned with our quality criteria with the mean effect size of the meta-analysis.

Preliminary Results

The corpus consists of 14 studies published between 2015 and 2025. One study in 2015, one in 2020, three in 2022, two in 2023, three in 2024, and four in 2025. Across all studies, the total number of participants amounted to 712.

The studies targeted a range of language learning domains. Vocabulary was the most frequently studied (7 of 14), contrasting with previous reviews where speaking skills were most common (e.g., J. Chen et al., 2022; Parmaxi, 2023).

None of the studies reported negative effects of VR on L2 outcomes. Most showed positive impacts, though some were more nuanced (Xie et al., 2025). Immersion levels across studies were evenly split between high and low. With respect to task type, eight studies featured story-driven tasks, and six used open-ended ones. Participants engaged with non-human characters in six studies, while in the remaining eight they did not.

Effect size calculations are underway and will estimate VR's impact on L2 performance, including potential moderating factors such as immersion or language skill. This analysis aims to clarify the overall effectiveness of VR for L2 learning and inform future research and practice.

References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Chen, B., Wang, Y., & Wang, L. (2022). The effects of virtual reality-assisted language learning: A meta-analysis. *Sustainability, 14*(6), Article 3147. https://doi.org/10.3390/su14063147

Chen, J., Dai, J., Zhu, K., & Xu, L. (2022). Effects of extended reality on language learning: A meta-analysis. *Frontiers in Psychology, 13,* Article 1016519. https://doi.org/10.3389/fpsyg.2022.1016519

Cheung, A. C. K., & Slavin, R. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283–292. http://doi.org/10.3102/0013189X16656615

Cowie, N., & Alizadeh, M. (2022). The affordances and challenges of virtual reality for language teaching. *International Journal of TESOL Studies, 4*(3), 50–65. https://doi.org/10.46451/ijts.2022.03.05

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends, 66*, 810–824. https://doi.org/10.1007/s11528-022-00717-w

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*(4), 341–370. https://doi.org/10.3102/1076998606298043

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta- regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Higgins, J. P. T., Savovic, J., Page, M. J., & Sterne, J. A. C. (2019). *Revised Cochrane Risk-of-Bias tool for randomized trials. (RoB 2)*. Retrieved October 16, 2025 from https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2

Haoming, L., & Wei, W. (2024). A systematic review on vocabulary learning in AR and VR gamification context. *Computers & Education: X Reality, 4*, Article 100057. https://doi.org/10.1016/j.cexr.2024.100057

Huang, H.L., Hwang, G.J., & Chang, C.Y. (2020). Learning to be a writer: A spherical video-based virtual reality approach to supporting descriptive article writing in high school Chinese courses. *British Journal of Educational Technology, 51*(4), 1386–1405. https://doi.org/10.1111/bjet.12893

Kaplan-Rakowski, R., & Gruber, A. (2019). Low-immersion versus high-immersion virtual reality: Definitions, classification, and examples with a foreign language focus. In *Proceedings of the Innovation in Language Learning International Conference 2019: Florence*. Pixel.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Sage.

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*, 937–958. https://doi.org/10.1007/s10648-020-09586-2

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/bm.2012.031

Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods, 11*(2), 364–386. http://doi.org/10.1177/1094428106291059

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S.,... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article 71. https://doi.org/10.1136/bmj.n71

Parmaxi, A. (2023). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments, 31*(1), 172–184. https://doi.org/10.1080/10494820.2020.1765392

Pustejovsky, J. (2020). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections*. https://CRAN.R- project.org/package=clubSandwich

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science, 23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Springer.

Slavin, R. E., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness, 4*(4), 370–380. http://doi.org/10.1080/19345747.2011.558986

Slavin, R. E., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Education Evaluation and Policy Analysis, 31*(4), 500–506. http://doi.org/10.3102/0162373709352369

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Xie, T., Zhang, H., & Yang, Y. (2025). Effect of immersive virtual reality based upon input processing model for second language vocabulary retention. *Education and Information Technologies, 30*(9), 12365–12385. https://doi.org/10.1007/s10639-025-13333-x

Yu, Z., & Xu, W. (2022). A meta-analysis and systematic review of the effect of virtual reality technology on users' learning outcomes. *Computer Applications in Engineering Education, 30*(5), 1470–1484. https://doi.org/10.1002/cae.22532