Magical World VR: Enhancing English Vocabulary Acquisition for Elementary Learners through Virtual Reality

Yuli Astutik[1,*], Nur Farhana bte Abd Wahid[2], Rahmania Sri Untari[3], and Fitria Wulandari[4]

[1]English Language Education, Universitas Muhammadiyah Sidoarjo, Sidoarjo, Indonesia
[2]Early Childhood Education, Universiti Tunku Abdul Rahman, Kajang, Malaysia
[3]Information Technology Education, Universitas Muhammadiyah Sidoarjo, Sidoarjo, Indonesia
[4]Teacher Professional Education, Universitas Muhammadiyah Sidoarjo, Sidoarjo, Indonesia
Email: yuliastutik@umsida.ac.id (Y.A.); farhanaw@utar.edu.my (N.F.B.A.W.); rahmania.sriuntari@umsida.ac.id (R.S.U.); fitriawulandari1@umsida.ac.id (F.W.)
*Corresponding author
Manuscript received April 27, 2025; revised June 4, 2025; accepted August 19, 2025; January 9, 2026

***Abstract*—This study aimed to design and examine a Virtual Reality (VR)-based instructional media to support English vocabulary learning among Indonesian elementary students. In response to the limited use of immersive technology in Southeast Asian primary classrooms, the study followed a Research and Development (R&D) approach, incorporating expert review through the Delphi method and implementation using a quasi-experimental design. The developed media featured a curriculum-aligned, garden-themed virtual environment intended to facilitate contextual and interactive learning experiences. Participants were 45 fourth-grade students assigned at the class level to either a VR group ($n = 22$) or a realia group ($n = 23$). Both groups showed significant vocabulary gains from pre- to post-test ($p < 0.05$); however, no statistically significant difference was found between the groups ($p = 0.728$). To complement the test results, qualitative data were collected from classroom observations and teacher reflection logs. Students in the VR group demonstrated observable patterns of increased participation, peer collaboration, and spontaneous use of target vocabulary throughout the sessions. These findings indicate that while VR may not have produced significantly higher test outcomes compared to traditional methods, it offered observable benefits related to learner behavior and engagement. The study presents a locally grounded, pedagogically informed model of VR use in early language learning. Further research is recommended to explore its long-term impacts and implementation across varied educational contexts.**

***Keywords*—virtual reality, vocabulary acquisition, elementary education, language learning, immersive media**

I. Introduction

English language instruction in elementary schools plays a vital role in shaping students' foundational communication skills [1–4]. As English continues to be the lingua franca of global education, commerce, and digital communication, early mastery of vocabulary becomes essential [5]. Traditional vocabulary teaching methods in primary education, however, often rely on rote memorization and passive learning techniques, which may hinder engagement and long-term retention [6, 7]. Recent educational research highlights the need for more immersive and contextualized learning environments to better support young learners' language development [8–10]. The use of technology in learning can help create a more interactive and enjoyable learning environment for students. One technology that has great potential to revolutionize the way learning and teaching are conducted is Virtual Reality (VR). VR technology allows the creation of realistic simulated environments, where students can interact with objects and situations that seem real. According to Lan and Predescu *et al.*, Virtual Reality (VR) is introduced as an innovative tool that has great potential in enhancing English language learning [11, 12]. With the use of realistic environmental simulations, Virtual Reality (VR) may produce a more captivating and immersive learning environment for students. Through the use of a virtual reality headset, students can immerse themselves in a relevant and engaging situation while learning vocabulary in English. Students can learn vocabulary by using virtual environments, such as plants, animals and other objects, and pronounce the words in appropriate contexts [13, 14].

While VR has been increasingly adopted in secondary and higher education [15–17], its use in primary level English instruction particularly for vocabulary acquisition remains limited and underexplored, especially in non-English-speaking countries like Indonesia. The lack of contextually relevant and empirically validated VR-based instructional media for young learners presents a critical gap in language education research. This study addresses this problem by developing and testing a VR-based learning tool designed specifically to support vocabulary learning among elementary school students.

II. Literature Review

English language instruction at the elementary level presents a number of persistent challenges. One major concern is the absence of interactive and engaging instructional media that can stimulate students' interest and deepen their comprehension of new vocabulary [18, 19]. Conventional approaches tend to be teacher-centered and passive, requiring students to merely listen and memorize lists of words, often devoid of meaningful context. This often results in diminished motivation and reduced learning effectiveness [7]. Moreover, the lack of contextualized and immersive learning environments may hinder students' ability to internalize vocabulary and apply it meaningfully in authentic situations. Consequently, there is an increasing demand for more dynamic, responsive, and context-rich instructional tools tailored for young learners.

A growing body of research has demonstrated the benefits of Virtual Reality (VR) in language learning, particularly in terms of enhancing learner motivation, engagement, and overall achievement [12, 20–22]. Lin and Lan [21] assert that VR holds substantial potential in the field of language education and is likely to evolve further alongside technological innovations. Similarly, Huang *et al.* [22] confirmed that both VR and Augmented Reality (AR) contribute positively not only to academic achievement but also to fostering more meaningful interaction between learners and learning content. However, these studies tend to emphasize generalized outcomes and often lack attention to the specific pedagogical needs of young learners. Aspects such as cognitive appropriateness, interface simplicity, and vocabulary scaffolding for children remain underexplored, thus limiting the direct applicability of their findings to early English language education.

Complementary findings from adjacent fields further reinforce VR's pedagogical value. Freina and Ott [20], whose work was situated outside language education, found that VR significantly enhances users' spatial orientation and immersive awareness, despite occasional sensory dissonance between virtual and physical environments. Lan [12] offers additional evidence that VR integration in learning can significantly boost motivation and outcomes, even among learners with Autism Spectrum Disorders (ASD). In his study, students demonstrated increased creativity, autonomy, and communicative competence within collaborative VR-based tasks. At the tertiary level, Young *et al.* [23] and Nicolaidou [24] examined that university students responded positively to VR-enhanced learning environments, reporting higher engagement, enjoyment, and perceived control, despite encountering occasional physical and technical challenges. While these findings highlight the broad cognitive and emotional benefits of VR, they largely pertain to non-primary and special-needs contexts and do not directly inform instructional design for early English language acquisition.

Recent studies involving young learners outside Southeast Asia also underscore the promise of immersive technologies in early language instruction. For instance, Korosidou [25] reported that Greek children aged 5–6 improved their vocabulary and motivation after using AR to learn the alphabet and contextual vocabulary. Dooly *et al.* [26], working with 10- to 11-year-old EFL students in Spain, found that VR environments encouraged spontaneous language use and increased learner confidence, even though minor decreases in grammatical accuracy were observed. Xie, Zhang, and Yang [27], in a study with fifth-grade Chinese students, discovered that VR was particularly effective for long-term vocabulary retention when compared to traditional PowerPoint-based instruction, especially when grounded in input processing theory. Although these studies are insightful, they are often limited by their cultural and curricular misalignment with Southeast Asian classroom contexts, particularly Indonesia. Moreover, few of them engage directly with national curriculum goals or linguistic environments similar to that of Indonesian elementary schools.

A meta-analysis conducted by Merchant *et al.* [28] lends further support to VR's educational value. Their studies indicate that VR has shown pedagogical promise in a variety of educational settings, though outcomes may vary depending on implementation and context, from primary to higher education, highlighting its ability to address longstanding shortcomings of traditional instruction such as lack of engagement and contextual relevance. Despite these promising findings, empirical studies that examine VR's implementation for English vocabulary learning in Indonesian elementary classrooms remain sparse.

The present study seeks to bridge this gap by designing, developing, and validating a VR-based instructional tool that aligns with Indonesia's elementary English curriculum. It specifically aims to explore how immersive VR environments can be used to enhance young learners' vocabulary acquisition through interactive, culturally contextual, and curriculum-relevant design. Accordingly, the present study is guided by the following research questions:

1) What are the consensus results achieved in the development process of VR-based learning media?
2) What are the main features that should be included in effective VR-based English learning media?
3) To what extent does the use of VR media affect elementary students' vocabulary acquisition?
4) How does VR media compare to Realia media in primary school students' vocabulary acquisition?

III. Materials and Methods

A. Research Design

This study employed a Research and Development (R&D) approach, incorporating the Delphi method and an embedded mixed methods design. The R&D approach was used to design, develop, and validate Virtual Reality (VR)-based learning media for English instruction at the elementary school level. The Delphi method was applied to gather expert consensus on the media's core features, design elements, and feasibility. To assess the instructional impact of the media, the researchers employed a quasi-experimental design involving intact classroom groups. Additionally, qualitative data from classroom observation checklists and teacher reflection logs were embedded within the experimental phase to support and enrich the interpretation of quantitative results. This design enabled a more comprehensive understanding of how the VR media influenced student engagement, interaction, and vocabulary use in authentic classroom settings.

B. Development and Validation Procedures

1) Development approach

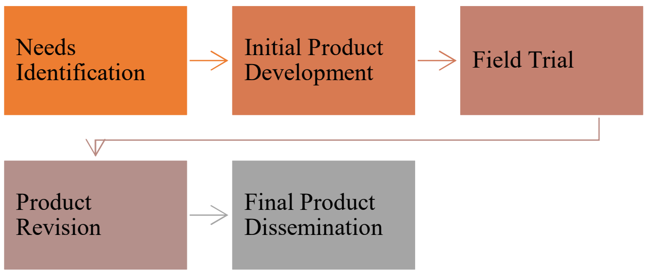

Fig. 1. The development process based on Borg and Gall [29].

The development process followed the stages outlined in Borg and Gall's R&D model [29], which was adapted into five essential stages, as shown in Fig. 1: (1) Needs Identification, which examined students' and teachers' requirements for vocabulary-supportive immersive media; (2) Initial Product Development, involving the creation of a VR prototype aligned with the curriculum and needs analysis; (3)

Field Trial, where the prototype was tested on a small group of non-sample students, with feedback collected from both learners and teachers; (4) Product Revision, which integrated improvements based on field trial results and expert input through the Delphi technique; and (5) Final Product Dissemination, where the validated media was implemented in real classroom settings for broader instructional use.

This adapted model retained the core structure of Borg and Gall's framework while adjusting it to the practical demands of classroom-based media development and validation.

To ensure the product's validity, relevance, and feasibility, the R&D process was further validated using the Delphi method, involving experts in English education, instructional media, and VR technology. Their input strengthened the overall development cycle and improved the instructional quality of the final VR product.

2) Delphi method and expert validation

a) Participants (expert panel)

The expert panel in this study consisted of six professionals with demonstrated expertise in education, Virtual Reality (VR) technology, and English language instruction. They were purposefully selected based on their academic qualifications, professional background, and scholarly contributions in relevant fields. Specifically, the panel included: (1) three education experts with extensive experience in elementary-level teaching and curriculum development; (2) two VR technology specialists with a track record in designing educational VR applications; and (3) one expert in English language learning with a strong background in instructional methodology and research related to early English education.

Although the panel was limited to six experts, this number falls within the commonly accepted range for Delphi studies, particularly in educational technology research. According to Hsu & Sandford, studies have shown that panels of 5–10 experts are sufficient when participants are well-qualified in their expertise [30].

b) Procedures and implementation

According to Laksmiwati *et al.* [31], Masnan *et al.* [32], and Toločka & Maceika [33], the Delphi method is used to gather expert views with the aim of reaching a consensus on instructional media.

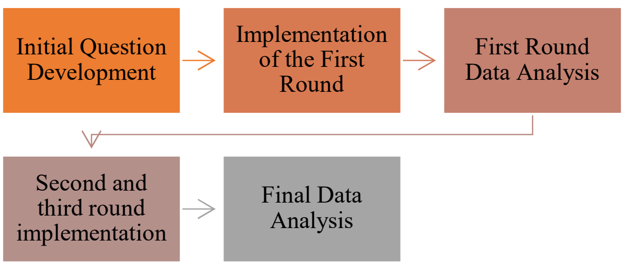

Fig. 2. The Delphi process [33].

In this study, the Delphi method was adapted from Toločka and Maceika [33] and implemented in multiple rounds to refine the features of the VR-based learning media. However, Unlike Toločka's approach, which used closed-ended questionnaires and statistical analysis, this study employed open-ended questionnaires analyzed through qualitative content analysis. This strategy allowed experts to provide in-depth feedback without constraints, enabling the researchers to identify recurring themes, categorize insights, and trace the development of consensus across rounds.

As illustrated in Fig. 2, the process began with the formulation of questions based on literature review and preliminary expert consultations. The first round involved distributing open-ended questionnaires to gather initial suggestions. Based on the responses, a second-round questionnaire was developed to verify and refine the identified themes. The process continued into a third round until expert feedback stabilized. Consensus was defined as the point where at least five out of six experts consistently supported the proposed features, with no new themes or substantial revisions emerging. This stability, combined with active participation from all panelists in each round, ensured the reliability and validity of the expert-driven evaluation.

C. Experimental Design and Classroom Implementation

1) Quasi experimental design

After validating the VR-based learning media through expert review and a small-scale trial involving 10 non-sample students, the researchers implemented a quasi-experimental design, specifically a non-equivalent group design. The experiment aimed to assess the impact of VR-based instruction on elementary students' English vocabulary acquisition and to compare it with the use of realia media.

The participants were fourth-grade students from Muhammadiyah Creative Elementary School in Bangil, East Java. There were four Grade 4 classes available at the time of the study. Based on input from the English teacher, two classes were recommended for inclusion in the study because they had consistently scored below the school's minimum passing grade in English and had never been exposed to VR-based instruction. These two intact classes had previously used conventional media such as realia, PowerPoint slides, pictures, and puppets.

From the two recommended classes, a random assignment was conducted at the class level using a random number generator to determine which class would serve as the experimental group (22 students) and which as the control group (23 students), resulting in a total of 45 student participants. Since randomization occurred at the class level rather than the individual level, this study does not meet the criteria for a true experimental design and is therefore classified as a quasi-experimental design.

This selection process limited the generalizability of the findings to the full population of Grade 4 students but allowed for the inclusion of learners who could benefit most from vocabulary support. The quasi-experimental nature of the study has been fully acknowledged in both the design and analysis, and efforts were made to minimize bias through procedural equivalence, consistent instruction, and standardized learning materials.

2) Classroom implementation procedures

The intervention spanned nine sessions of 70 minutes each, conducted from May to July 2024. Both groups followed the same schedule and content structure. Initial sessions introduced the learning media and familiarized students with garden-themed vocabulary (e.g., trees, flowers, grass), followed by thematic lessons covering plants, animals, and tools (e.g., sunflower, butterfly, watering can). Activities included games, role-plays, sentence construction, and Q&A

tasks. A pre-test was administered prior to the first session to establish baseline vocabulary knowledge, followed by a post-test after the final session to measure learning gains. The intervention focused on 25 target vocabulary items included: flower, butterfly, bee, tree, grass, watering can, soil, seed, leaf, shovel, branch, garden, fruit, root, stem, pot, bush, insect, vine, ant, fly, trunk, shoot, sprout, and garden path.

The number and duration of sessions were aligned with the national curriculum requirements for fourth-grade English instruction at the research site, which emphasize mastery of environmental vocabulary. Furthermore, all students participated from both classes consistently throughout the program, ensuring equal exposure to the instructional interventions across both groups.

To ensure procedural equivalence and minimize bias, both groups were taught by the same teacher, who received prior training to implement the lesson plans consistently. All students participated throughout the intervention period, ensuring equal exposure across both conditions.

- Ethical Consideration

Ethical and safety protocols were established before implementation. Informed consent was obtained from all guardians, and approval was granted by school authorities. An orientation session was held to familiarize students with the VR headset. During each session, students were continuously supervised by the teacher and research team. Participants were allowed to remove the headset at any time in case of discomfort. No incidents of cybersickness or distress were reported. This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. While no formal IRB approval was obtained due to institutional limitations, the research procedures strictly adhered to ethical standards, including informed consent, participant safety, and respect for participant autonomy.

3) Instrument for experimental design

In addition to the learning media (VR and realia), lesson plans and instructional materials were developed and validated by expert reviewers. To assess learning outcomes, pre-test and post-test instruments were designed in alignment with the Grade 4 English curriculum, specifically focusing on vocabulary related to environmental themes. Each test comprised 20 multiple-choice items and 5 word-to-picture matching items, with equivalent content and difficulty levels maintained across both versions. No formal oral test was administered; however, pronunciation development was tracked qualitatively via classroom observation and teacher reflections.

The test instruments were reviewed for content validity by three experts one English education lecturer and two experienced elementary school English teachers. A pilot test was subsequently conducted with 15 students from a different school to evaluate the quality of the items empirically. Item analysis was used to revise questions based on their clarity, difficulty, and discrimination power. Reliability analysis using Cronbach's alpha produced a coefficient of 0.812, indicating a high level of consistency (see Table A1). Based on these results, the instruments were considered valid, reliable, and ready for implementation in the experimental phase.

No a priori power analysis was conducted prior to data collection due to institutional limitations. The final sample size ($n = 45$) reflected full classroom participation.

D. Observation and Teacher Reflection

To triangulate the results of the true experimental design, this study also used qualitative data sources in the form of classroom observations and teacher reflection logs. These instruments were employed not to measure learning outcomes directly, but to monitor the fidelity of implementation and support the interpretation of quantitative findings.

Classroom observations were guided by structured checklists that captured student behaviors during each intervention session. Key observation indicators adapted from Candee included engagement (e.g., maintaining focus, showing enthusiasm), interaction (e.g., responding to instructions, collaborating with peers), and language use (e.g., recognizing and pronouncing vocabulary correctly) [34]. Two researchers conducted the observations to ensure consistency, marking each behavior as observed or not observed across the sessions.

In addition, teacher reflection logs were completed at the end of every session. These logs consisted of open-ended narratives describing the teaching process, students' reactions, challenges encountered, and suggestions for future improvements. The combination of observation checklists and teacher reflections functioned as process validation tools, helping to ensure that instructional delivery remained aligned with the planned procedures and learning objectives.

E. Data Analysis

This study employed both qualitative and quantitative data analysis procedures to address the research objectives comprehensively. For the development and validation of the VR-based English learning media, qualitative analysis was conducted using the Delphi method. Experts' responses from open-ended questionnaires were analyzed through content analysis. The researchers coded the data manually, grouped similar ideas into categories, and identified recurring themes across rounds. Thematic saturation and response stability were used to determine consensus, defined as consistent agreement among at least five out of six experts with no significant new input emerging in the final round. These insights guided the iterative refinement of the media features and instructional design.

To compare learning outcomes between the VR and realia groups, quantitative analysis was conducted using SPSS 25. First, the pre-test and post-test data were tested for normality using the Kolmogorov-Smirnov and Shapiro-Wilk tests (Table A2). Since the assumptions for parametric testing were met, a paired samples t-test was applied to compare vocabulary learning gains between the experimental and control groups. Then, the independent t-test was used to determine whether the difference in gains was statistically significant, thus assessing the impact of VR-based instruction compared to realia-based instruction. No missing data were found in either group.

In addition, qualitative were summarized descriptively showing students' behaviors and utterances observed across the nine sessions. These data were tabulated per session to assess trends in classroom behavior following the aspects adapted from Candee [34]. Meanwhile, teacher reflection logs were also descriptively analyzed to extract recurring narratives related to student participation, instructional challenges, and classroom dynamics. Teacher reflection logs were analyzed using inductive thematic coding by two researchers. Interrater reliability was established through consensus discussion. The reporting of this study follows key elements of the APA Journal Article Reporting Standards (JARS–Quantitative and Qualitative) as outlined by Appelbaum *et al.* [35] and Levitt *et al.* [36].

IV. Results and Discussion

A. RQ1: What Are the Consensus Results Achieved in the Development Process of VR-Based Learning Media?

As this study employed open-ended questionnaires in the Delphi process, responses were analyzed descriptively using content analysis. Thus, quantitative metrics such as average ratings were not applicable. The level of expert agreement was instead indicated by consistency of responses across rounds. The following is a detailed explanation the results of each Delphi round (Table 1):

Table 1. Delphi analysis of the development of VR media for elementary schools

Questionnaire Distribution	First Round	Second Round	Third Round
Number of Panellist	6 Expert Panel	6 Expert Panel	6 Expert Panel
Response Distribution	The majority of experts agreed on the importance of interactive features and realistic simulations in VR-based learning media. There were variations in opinions on certain technical details such as content difficulty level and user interface.	Consensus began to form with 5 experts agreeing on the proposed list of key features. Differences in opinion decreased, particularly after clarification and revision of the initial recommendations.	Consensus was nearly achieved, with 5 experts agreeing on all proposed features and 1 expert providing a minor suggestion for improvements were incorporated into the final media design.
Key Findings	The importance of interactivity and realistic simulation is considered crucial. Variations in opinions on content difficulty and interface.	The list of key features was almost entirely agreed upon. Clarifications helped reduce differences in opinion.	High consensus on the proposed features. Minor suggestions were added to refine the design.
Data Analysis	Content analysis of the initial responses highlighted a focus on interactive features and simulations. Variations in technical details were identified as areas for further improvement.	Data analysis showed progress toward consensus, with many experts agreeing on the key features. Changes based on first-round feedback were considered.	Analysis showed that almost all experts agreed with the final features. Minor suggestions were accepted and integrated into the final design.

The Delphi findings indicated a clear movement toward consensus. In the first round, the experts provided varied responses, especially regarding content scope and interface complexity. By the second round, five out of six experts supported the refined list of features. In the third round, five out of six experts maintained consistent feedback, and only one expert provided a minor wording suggestion, and no new themes emerged. This confirmed that responses had stabilized and consensus had been achieved (Table 1).

Table 2. Features of VR- based learning media

No.	Main Features	Sub Features	Function
1	High Interactivity	Logo	Introduces the VR environment and sets the theme for the learning experience.
		Home Screen	Provides the initial interface, guiding students on how to navigate the VR environment.
		Pointer	Allows students to interact with objects by pointing, facilitating hands-on engagement.
		Direction	Guides students within the VR space, ensuring smooth navigation through the scenes.
		Movement	Enables students to move within the VR environment, enhancing immersion and exploration.
2	Realistic Simulation	Garden Environment	Mimics a real-world garden setting, providing a familiar and relatable context for learning.
3	Quick Feedback	Popup Text on Objects	Provides immediate feedback by displaying the English name of each object when interacted with.
4	Relevant Content	Garden Vocabulary (e.g., flower, butterfly)	Aligns with the elementary English curriculum, teaching relevant vocabulary in context.
5	Usability	Intuitive User Interface	Facilitates ease of use for both students and teachers in navigating the VR media.

Following the consensus-building results about VR media used in primary school English language instruction, the researchers then focus towards outlining the key features of this media.

B. RQ2: What Are the Main Features That Should be Included in Effective VR-Based English Learning Media?

The consensus results from the expert panel include several key features that must be present in VR-based learning media is explained in Table 2.

Experts agreed on several core features that can enhance the usability and instructional potential of VR-based English learning media for young learners. These include high interactivity, realistic simulation, immediate feedback, curriculum-aligned content, and usability (Table 2).

To illustrate the practical implementation of these features, the "Magical World VR" application integrates multiple interactive components designed specifically for young learners. Upon launching the application, students are welcomed with a home screen featuring a virtual school entrance and tutorial instructions for navigation (Fig. 3). A central pointer allows students to interact with virtual objects by aligning their gaze with 3D items. This intuitive feature supports hands-free engagement and reduces cognitive load. Navigation within the virtual garden is controlled through simple head gestures, enabling learners to explore the environment naturally without requiring handheld controllers.

Within the garden-themed VR environment, each object such as flowers, insects, or tools is embedded with contextual vocabulary. When students direct the pointer toward an object, a popup label appears instantly, displaying the English word associated with it (Fig. 4). This immediate feedback reinforces vocabulary recognition and facilitates real-time language processing.

Fig. 3. Home screen of “Magical World VR”, featuring the school entrance and tutorial instructions for navigation.

(a)
(b)
Fig. 4. Pointer interaction and popup feedback. When learners direct their gaze to an object, a label appears showing the English vocabulary word.

The setting itself is designed to closely simulate a real-world garden with lifelike textures and sounds, fostering contextual learning through spatial immersion (Fig. 5). All content is aligned with Indonesia’s national curriculum for fourth-grade English, focusing on environmental vocabulary. Moreover, the interface prioritizes ease of use, ensuring accessibility for both students and teachers with varying levels of technological familiarity.

Together, these integrated features contribute to a pedagogically sound and learner-centered VR learning experience. The combination of immersive simulation, responsive interactivity, and curriculum relevance assist not only vocabulary acquisition but also engagement, motivation, and meaningful language use.

Fig. 5. Realistic garden simulation representing the immersive learning environment aligned with environmental vocabulary content.

C. RQ3: To What Extent Does the Use of VR Media Affect Elementary Students' Vocabulary Acquisition?

The statistical assumptions for parametric testing were met, as confirmed by the Kolmogorov-Smirnov and Shapiro-Wilk tests for normality and Levene’s Test for homogeneity of variance ($p > 0.05$), as presented in Tables A1 and A2. Specifically, Levene’s Test indicated that the variances of vocabulary gain scores between the VR and realia groups were homogeneous (Levene’s statistic = 0.013, p = 0.910), as presented in Table A3. Based on these results, a paired samples t-test was conducted to examine vocabulary score differences before and after the intervention in both the experimental (VR) and control (realia) groups.

As shown in Table 3, both groups exhibited statistically significant improvements in vocabulary scores. The experimental group (VR) showed a mean difference of −10.909 ($t(21) = -2.347$, $p = 0.029$), while the control group (realia) had a mean difference of −9.565 ($t(22) = -2.080$, $p = 0.049$). Thus, indicate positive learning gains. Moreover, the VR group demonstrated a slightly greater average improvement than the control group.

To assess the practical impact, effect size calculations were included. The VR group achieved a medium effect size (Cohen’s $d = 0.55$), whereas the realia group yielded a small-to-medium effect size ($d = 0.50$).

These findings are further visualized in Fig. 6, which illustrates the pre- and post-test vocabulary scores in both groups. While the statistical analysis revealed no significant difference, the visual trend and classroom observations suggest that VR media may support increased learner engagement and experiential learning.

Table 3. Paired samples test (Paired Differences)

Pair	Comparison	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Pre Test Experiment - Post Test Experiment	−10.909	21.802	4.648	−20.576	−1.243	−2.347	21	0.029
Pair 2	Pre Test Control - Post Test Control	−9.565	22.049	4.598	−19.100	−0.030	−2.080	22	0.049

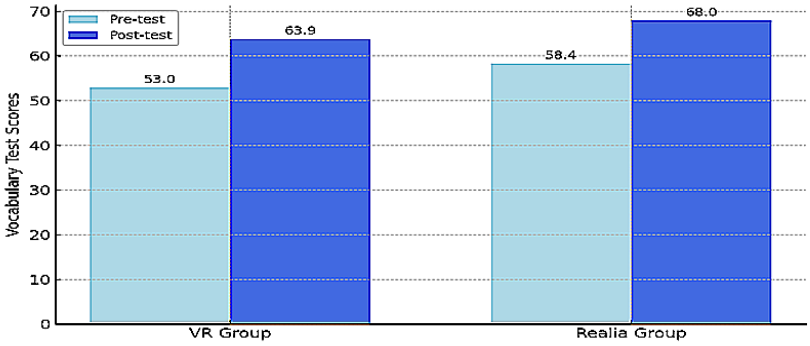

Fig. 6. Pretest and postest vocabulary scores in VR and realia group.

D. RQ4: How Does VR Media Compare to Realia Media in Primary School Students' Vocabulary Acquisition

To address RQ4, an independent samples t-test was conducted to compare the post-test vocabulary scores between students who received instruction using VR media and those who were taught using realia-based materials. Prior to the comparison, Levene's Test for equality of variances was performed to ensure the homogeneity assumption was met. As shown in Table 4, the result ($F = 0.013$, $p = 0.910$) indicated that the variances between the two groups were equal, justifying the use of the t-test for further analysis.

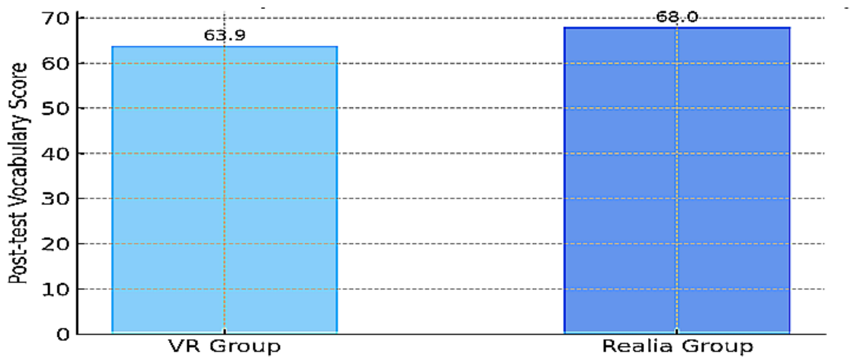

Fig. 7. Posttest scores comparison between VR and realia groups.

Table 4. Independent samples test (Students test result)

Equality of Variances	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	0.013	0.910	0.350	43	0.728	1.719	4.916	−8.195	11.634
Equal variances not assumed			0.349	42.742	0.728	1.719	4.920	−8.204	11.643

The independent samples t-test revealed no statistically significant difference in post-test vocabulary scores between the VR and Realia groups, $t(43) = 0.350$, $p = 0.728$. Although both instructional methods led to vocabulary improvements (as shown in RQ3), the post-test performance between groups was comparable, suggesting that VR did not yield greater learning gains than realia-based instruction in terms of test outcomes.

To interpret the practical magnitude of this difference, the effect size was calculated, yielding Cohen's $d = 0.10$, classified as negligible. This result aligns with the statistical outcome and is illustrated in Fig. 7, which compares post-test means across groups.

While the immersive nature of the VR experience elicited greater behavioral engagement, this did not translate into significantly higher test scores during the short intervention period. Moreover, test-based assessments may not fully capture the depth of learning that occurs through interactive media.

Given the small observed effect size, a post hoc power analysis was conducted, yielding an estimated power of only 5.4%. This indicates that the study was underpowered to detect small effects. A hypothetical a priori power analysis further showed that approximately 788 participants (394 per group) would be required to detect such a small effect with 80% power at $\alpha = 0.05$.

These findings highlight the need for cautious interpretation of the results and underscore the importance of considering complementary outcome measures such as student behavior, motivation, and spontaneous language use as discussed in the following qualitative analysis.

- Observation and Teacher Reflection Results

To complement the quantitative findings, classroom observation and teacher reflection data were analyzed to assess student engagement, interaction, and vocabulary use during the VR-based instruction sessions. These data offer additional insights, particularly in light of the statistical analysis from RQ4, which showed no significant difference between the VR and realia groups. Despite comparable test scores, students in the VR group consistently demonstrated higher levels of behavioral engagement and language production, which may not be fully captured through quantitative measures alone.

Table 5. Observation checklist summary on students in the implementation of "Magical World VR"

Session	Engagement (Focus & Enthusiasm)	Interaction (Peer & Teacher Response)	Language Use (Vocabulary Identification & Pronunciation)	Examples of Student Actions / Utterances
M1	Students were curious but distracted. Some hesitated to wear the headset.	Minimal peer talk; mostly silent or passive responses.	Students struggled to say words like "tree" or "flower."	One student asked, "*Bu, ini taman beneran ya?*" (Ms, is this a real park?) Another pointed and said, *"Itu pohon ya, Bu?*" (Is that a tree, Ms.?)
M2	More attentive during object exploration; excited when hearing pop-up sounds.	Students began asking questions to the teacher about what to click.	Started repeating basic words like "tree", "grass", and "bee" after hearing them.	Student said, *"Kalau saya lihat lebah, bunyinya muncul ya?"* (When I see a bee, it makes a sound?) and repeated "bee, bee, bee" several times.
M3	Sustained attention improved. Students navigated VR with less assistance.	Students began sharing what they saw with friends, "*yayy, aku ke kebun belakang sekolah*" (Look, I am in the garden behind the school)	Pronunciation slightly improved; students mimicked words from the VR pop-ups.	A student practiced saying "sunflower" and then asked, "*Bu, ini bunga matahari?"*(Ms. Is this a sunflower?)
M4	Students were excited and waited their turn more patiently.	Some students helped others adjust their head position.	Recognized words faster, especially when seeing garden animals.	*"Tadi aku lihat kupu-kupu… hmm Butterfly kan, Bu?"* (I saw a butterfly….hmm Butterfly, right Ms?) said a student to the teacher.
M5	Students engaged more during game-based vocabulary activities.	Encouraged friends to say the words aloud.	Used new words in short phrases: "This is a bee", "That is a tree."	*"Aku tahu 'watering can', itu buat nyiram, kan?"*("I know 'watering can', it's for watering, right?")
M6	Very focused on tasks; enjoyed finding hidden garden objects.	Interactive Q&A became more natural.	Spontaneous use of English words during play.	"Where is the flower?" one student asked while pointing at a 3D object.
M7	Enthusiasm remained high; students explored deeper into the virtual garden.	Students initiated group tasks and shared VR tips.	Began using full sentences like "I see a shovel."	Student shouted excitedly, "*Aku tahu aku tahu ... a butterfly flies in the garden is beautiful..." (*"I know I know... a butterfly flies in the garden is beautiful...")
M8	Students demonstrated confidence; some explored independently without teacher prompts.	Less reliance on teacher; peer interaction increased.	Students tried connecting vocabulary into written sentences.	One wrote "I like flower and grass" after the session.
M9	High engagement maintained until the end; students expressed sadness it was the last session.	Encouraged classmates during the final review game.	Used nearly all target vocabulary with better accuracy and pronunciation.	"*Bu, besok bisa pakai VR lagi nggak? Aku suka belajar begini*." (Ms. can you use VR again tomorrow? I like learning like this)

Note: M1–M9 refer to Meeting 1–Meeting 9.

Table 5 summarizes student development over nine sessions. In the early stage (M1), students showed curiosity but were distracted and hesitant to use the VR headset. Peer interaction was minimal, and most students struggled to produce even simple words such as "tree" and "flower." By the third session (M3), attention and navigation improved, and students began verbally responding to the virtual environment, repeating vocabulary heard from pop-ups, and engaging peers with comments like "*yayy, aku ke kebun belakang sekolah*" (I'm in the garden behind the school). In mid-sessions (M4–M6), students actively used new vocabulary in short phrases, engaged in peer support, and initiated interactive Q&A in English, such as "Where is the flower?"

In the final sessions (M7–M9), students explored more independently and confidently. They used full English sentences, helped peers without teacher prompting, and even attempted written language production. A student wrote "I like flower and grass" independently after a session, and another expressed enthusiasm by asking, "*Bu, besok bisa pakai VR lagi nggak? Aku suka belajar begini*" (Can we use VR again tomorrow? I like learning like this).

Teacher reflection logs further corroborated the observation data, providing descriptive insights into students' behavioral and linguistic development across sessions. In the earlier sessions, teachers noted students' hesitancy in operating the VR headset and their limited vocabulary output. However, across subsequent meetings, students showed growing independence, collaboration, and confidence in using English spontaneously. The following excerpts illustrate this progression.

Session 3 (M3):

"Today I see that my students no longer ask for help every time they navigate VR. They started to help each other operate the headset and pointer. Some of them dared to mention the word 'sunflower' without hesitation, indicating an increase in confidence in the use of vocabulary".

Session 6 (M6):

"The question-and-answer interaction became more natural. One student suddenly asked in English, 'Where is the flower?', without being prompted, and his friends answered using the vocabulary they had learned. This shows the students are really engaged and using the language spontaneously".

Session 9 (M9):

"They were enthusiastic about helping their friends when playing the review game, and almost all the target words were spoken with better accuracy. One student even asked, 'Ms. can we use VR again tomorrow? I like learning like using this (VR)'".

From the samples of teacher reflection (M3, M6, M9), indicate a clear shift from teacher-led to student-initiated learning. By the middle of the intervention period, students were not only navigating the VR interface with minimal assistance, but also demonstrating increased autonomy in using the target language. Vocabulary production became more confident and embedded in authentic communication, especially during peer interactions and game-based reviews. By the final session, students were observed supporting one another, producing full sentences, and engaging emotionally with the learning process.

This behavioral and linguistic progression, consistently recorded across teacher logs and observation checklists, reinforces the conclusion that VR-based instruction supports not only vocabulary acquisition, but also student motivation, collaboration, and contextual language use. These outcomes, although not fully captured in post-test scores, provide strong evidence of VR's pedagogical value in early language learning.

The study confirmed that the VR-based English learning media developed through the R&D process met expert-validated criteria for content, design, and usability, as assessed through the Delphi method. Classroom implementation revealed signs of student engagement, more active vocabulary use, and consistent behavioural indicators related to motivation and participation. While these patterns emerged from classroom observations and teacher reflections, they were not directly reflected in significantly test scores. Therefore, these findings should be interpreted as indicators of potential benefits rather than conclusive evidence of effectiveness.

The findings of this study align with those of several researchers who support the integration of immersive technologies in language learning. Studies by Lan [12] and Huang *et al.* [22] have demonstrated that virtual environments foster interaction and learner autonomy, particularly in contexts where language learning is supported by visual and situational cues. Similar to these studies, the present research found that students responded positively to the immersive features of the VR environment, such as gaze-based interaction and thematic navigation. These features encouraged spontaneous vocabulary production and increased learner confidence. The affective and behavioral benefits observed in this study reflect those noted by Dooly *et al.* [26] and Nicolaidou *et al.* [24], underscoring VR's potential to assist learner-centered, contextualized instruction.

This study shares similarities with Nicolaidou *et al.* [24], who found no significant difference in language learning outcomes between VR and mobile applications among higher education students. However, while their study emphasized user satisfaction and task involvement, the present study highlights spontaneous language use, peer interaction, and contextual vocabulary application among primary school learners in a curriculum-based setting. These process-based gains, though not fully captured in standardized test scores, suggest that VR may offer unique pedagogical value in early language education.

Compared to studies by Dooly *et al.* [26] and Korosidou [25], which explored VR and AR use among European elementary learners, the current research reinforces the finding that immersive environments can stimulate communicative confidence and vocabulary acquisition in children. However, by embedding VR within the structured scope of Indonesia's environmental vocabulary unit, this study advances the understanding of how localized, thematic content enhances the cognitive and emotional learning process. Related findings have been reported by Young *et al.* [23] and Freina and Ott [20], who emphasized VR's potential to deepen immersion and increase spontaneous interaction among learners.

These findings align with principles of constructivist learning theory, which emphasizes the importance of active participation and contextual engagement in supporting language development. The VR design in this study featuring object interaction, head-motion navigation, and sensory-rich environments enabled vocabulary learning to occur in meaningful contexts. This aligns with Huang *et al.* [22], and Merchant *et al.* [28], who showed that immersive technologies address traditional limitations in language instruction by promoting deeper engagement and contextual understanding.

From an implementation perspective, this study suggests that VR holds potential for use in English language learning at the elementary level, including in resource-limited settings. This potential can be realized if the technology is intuitive, aligned with the curriculum, and supported by adequate teacher preparation. Features such as head-motion navigation and simplified user interfaces may reduce technical barriers and support learner autonomy. Positive student responses reflected in enthusiasm, collaboration, and increased confidence suggest that VR is a promising and feasible instructional medium for primary-level language education.

Despite its contributions, this study has several limitations. Although the intervention was conducted over nine sessions, the post-test vocabulary scores remained modest in absolute terms. This outcome may be attributed to the contextual nature of the vocabulary, the novelty of the VR experience, and the gap between immersive instruction and conventional assessment formats. While the students demonstrated active engagement and contextual understanding during the sessions, these gains may not have been fully captured by the written vocabulary test. This limitation highlights the need for multimodal assessment strategies that better align with the dynamic learning processes enabled by VR-based instruction. Second, it was conducted in a single school with a relatively small sample size, limiting generalizability. This is consistent with what Nicolaidou *et al.* [24] also noted in their study, where contextual constraints influenced VR outcomes. In addition, this study measured only short-term vocabulary acquisition and did not assess long-term retention or transfer to productive skills, as explored in Xie *et al.* [27].

Another limitation relates to infrastructure. although this study did not face hardware or connectivity issues, such limitations may exist in other Indonesian schools, especially in more remote areas. Radianti *et al.* [37] and Atabek [38] have shown that technology integration success in education is often influenced by institutional readiness and infrastructure availability. Therefore, findings from this study may not be adopted in less-equipped environments without additional support. Additionally, while the VR sessions were well-tolerated by all students in this study, the ethical considerations of using immersive technology with young learners should not be overlooked. Safety measures, including guardian consent, headset orientation, and real-time supervision, were implemented to ensure a secure and comfortable learning experience.

Future research should explore the long-term effects of VR on vocabulary retention, speaking fluency, and integration into broader language tasks such as writing. Comparative studies involving augmented and mixed reality technologies, as seen in Robbani *et al.* [12] and Al-Said *et al.* [13] could reveal the unique strengths of each modality. Additional research should also examine scalable models for implementing VR in rural and low-connectivity schools. Insights from Halimah *et al.* [39] and Kurniawati *et al.* [40] offer promising frameworks for teacher-led innovation and curriculum integration. Finally, VR's application should be explored beyond language learning. Studies by Simonetti *et al.* [41] and Parong & Mayer [42] have demonstrated VR's potential in science and mathematics education, suggesting its interdisciplinary potential in K12 learning.

Although statistical analysis did not reveal a significant difference in vocabulary test scores between the VR and realia groups (RQ4), qualitative data such as classroom observations and teacher reflections suggest that VR may offer meaningful pedagogical benefits. In particular, the VR group demonstrated higher levels of engagement, motivation, and contextual language use, which align with constructivist and experiential learning principles.

The VR media developed in this study was validated by expert reviews and found to be practical during classroom implementation. Although its measurable impact on vocabulary acquisition was not statistically significant, students' behavioral and affective responses indicate that VR holds promise as a complementary instructional tool in elementary English learning.

Based on the post hoc power analysis, the statistical power was found to be only 5.4%, which is considered very low. This reflects a high risk of Type II error, likely due to the small sample size and minimal mean difference between the groups (i.e., low effect size). This result aligns with previous findings by Cohen [43] and Button *et al.* [44], who noted that underpowered studies especially those with small sample sizes and high variability tend to miss true effects. Given that participants were drawn from intact classes recommended by teachers (nonrandomized), the researchers had limited control over increasing the sample size. These findings emphasize the need for cautious interpretation of non-significant results and point to the necessity of stronger designs and larger samples in future research.

V. Conclusion

In conclusion, this study has explored the development and implementation of Virtual Reality (VR)-based media as a potential tool to support English vocabulary acquisition in elementary education. The VR media was systematically developed and validated, and was deemed feasible and relevant for classroom use based on expert evaluations and field trials. Although statistical analysis did not reveal significant differences in vocabulary test scores between the VR and realia groups, qualitative findings such as increased student engagement, spontaneous language use, and emerging motivation indicate promising pedagogical value of VR, particularly in fostering contextual and learner-centered language learning experiences.

Given the short duration of the intervention and the limited sample size, these findings should be interpreted with caution. Nevertheless, the results offer preliminary evidence that VR-based instruction has the potential to enhance the quality of language learning experiences. Future research is recommended to explore the long-term effects and assess its scalability across diverse educational contexts, particularly in the development of learner-centered instructional media innovations.

Appendix

Table A1. Item analysis of vocabulary test

No.	Item	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	Item 1	0.42	0.803
2	Item 2	0.45	0.802
3	Item 3	0.38	0.804
4	Item 4	0.51	0.800
5	Item 5	0.49	0.801
6	Item 6	0.44	0.802
7	Item 7	0.36	0.805
8	Item 8	0.53	0.799
9	Item 9	0.48	0.801
10	Item10	0.40	0.804
...	...	...	...
25	Item25	0.42	0.803

Note: All items showed acceptable corrected item-total correlations (≥ 0.36), and the reliability coefficient of 0.812 indicates high internal consistency

Table A2. Normality tests for vocabulary test scores (Student test results)

Class	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pre Test Experiment (VR)	0.168	22	0.110	0.915	22	0.060
Post Test Experiment (VR)	0.148	22	0.200*	0.915	22	0.061
Pre Test Control (Realia)	0.175	23	0.067	0.934	23	0.135
Post Test Control (Realia)	0.163	23	0.113	0.918	23	0.061

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction.

Table A3. Test of homogeneity of variance (Students test result)

Levene's Test	Levene Statistic	df1	df2	Sig.
Based on Mean	0.013	1	43	0.910
Based on Median	0.011	1	43	0.916
Based on Median and with adjusted df	0.011	1	42.421	0.916
Based on trimmed mean	0.010	1	43	0.920

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Y.A. contributes to creating research ideas, analysing data, and writing the article; N.F.B.A.W. contributes to validating research instruments and proofreading the article; R.S.U. contributes to creating the Virtual reality; F.W. contributes to collecting data, and helping Y.A. in analysing data; all authors had approved the final version.

ACKNOWLEDGEMENT

The researchers extend their heartfelt gratitude to all co-authors, respondents, participants, reviewers and journal editors who contributed to the success of this research and development project. Additionally, sincere thanks are given to the Directorate of Research and Community Service (DRPM) at Universitas Muhammadiyah Sidoarjo for generously funding and supporting the implementation of this research.

REFERENCE

[1] Y. Astutik and O. Purwati, "Verbal and nonverbal language: Pre-service teachers' decisiveness in TEYL," *J. Pendidik. Bhs. dan Sastra*, vol. 21, no. 1, pp. 1–12, Jul. 2021. doi: 10.17509/bs_jpbsp.v21i1.36651

[2] Y. Astutik, F. Megawati, and C. N. Aulina, "Total Physical Response (TPR): How is it used to teach EFL young earners?" *Int. J. Learn. Teach. Educ. Res.*, vol. 18, no. 1, pp. 92–103, 2019. doi: 10.26803/ijlter.18.1.7

[3] D. Kaur and A. A. Aziz, "The use of language game in enhancing students' speaking skills," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 10, no. 12, 2020. doi: 10.6007/ijarbss/v10-i12/8369

[4] Y. Astutik, S. Setiawan, S. Anam, and Suhartono, "'I can teach with my videos': How do teachers teach English to young learners in a technology-limited environment?" *Int. J. Learn. Teach. Educ. Res.*, vol. 21, no. 7, pp. 158–177, 2022. doi: 10.26803/ijlter.21.7.9

[5] S. Alhasnawi, "English as an academic lingua franca: Discourse hybridity and meaning multiplicity in an international Anglophone HE institution," *J. English as a Ling. Fr.*, vol. 10, no. 1, pp. 31–58, 2021. doi: 10.1515/jelf-2021-2054

[6] O. Ullah and M. Iqbal, "Comparison of impact of traditional and modern teaching methods on students' performance at elementary school level," *Glob. Reg. Rev.*, vol. 5, no. 1, pp. 386–395, 2020. doi: 10.31703/grr.2020(v-i).42

[7] M. A. Nguyen, T. H. Hanh, N. Thi, and H. Hanh, "the effectiveness of learning vocabulary through games in comparison with that in traditional technique," *Int. J. Educ. Res.*, vol. 04, no. 05, pp. 68–82, 2021.

[8] M. Altun and H. K. Ahmad, "The use of technology in English language teaching: A literature review," *Int. J. Soc. Sci. Educ. Stud.*, vol. 8, no. 1, 2021. doi: 10.23918/ijsses.v8i1p226

[9] R. P. D. Redondo, M. C. Rodríguez, J. J. L. Escobar, and A. F. Vilas, "Integrating micro-learning content in traditional e-learning platforms," *Multimed. Tools Appl.*, vol. 80, no. 2, pp. 3121–3151, 2021. doi: 10.1007/s11042-020-09523-z

[10] Z. Yang, "Application and exploration of VR and AI technology in college English teaching," *Adv. Multimed.*, vol. 2022, 2022. doi: 10.1155/2022/1810177

[11] S. L. Predescu, S. I. Caramihai, and M. A. Moisescu, "Impact of VR application in an academic context," *Appl. Sci.*, vol. 13, no. 8, 2023. doi: 10.3390/app13084748

[12] Y.-J. Lan, "Immersion into virtual reality for language learning," *Psychol. Learn. Motiv.—Adv. Res. Theory*, vol. 72, pp. 1–26, 2020. doi: 10.1016/bs.plm.2020.03.001

[13] A. S. Robbani, U. Baroroh, T. Musthofa, S. Purnama, S. Sugiyono, and A. Z. Annafiri, "The role of technology in language immersion: A systematic literature review," *Int. J. Eval. Res. Educ.* , vol. 13, no. 2, pp. 705–713, 2024. doi: 10.11591/ijere.v13i2.26733

[14] K. Al-Said, A. Berestova, N. Ismailova, and N. Pronkin, "The impact of video-based virtual reality training on critical thinking and cognitive load," *Int. J. Eval. Res. Educ.*, vol. 13, no. 5, p. 3239, 2024. doi: 10.11591/ijere.v13i5.28109

[15] A. Al-Gindy, C. Felix, A. Ahmed, A. Matoug, and M. Alkhidir, "Virtual reality: Development of an integrated learning environment for education," *Int. J. Inf. Educ. Technol.*, vol. 10, no. 3, pp. 171–175, 2020. doi: 10.18178/ijiet.2020.10.3.1358

[16] H. Huang and S. Liaw, "Applying situated learning in a virtual reality system to enhance learning motivation," *Int. J. Inf. Educ. Technol.*, vol. 1, no. 4, pp. 298–302, 2011. doi: 10.7763/ijiet.2011.v1.48

[17] M. Y. An, K. A. Ko, and E. J. Kang, "Problems and directions of development through analysis of virtual reality-based education in korea," *Int. J. Inf. Educ. Technol.*, vol. 10, no. 8, pp. 552–556, 2020. doi: 10.18178/ijiet.2020.10.8.1423

[18] K. Andre, K. Dewantara, N. Luh, P. Sri, and P. Suarcaya, "Trilingual supplementary learning material : Needs of English learning materials for first grade elementary school students," vol. 8, no. 2, pp. 287–293, 2024.

[19] X. Hu, "Research on assisted teaching based on VR—Taking Chinese school English course as an example," *SCIREA J. Educ.*, vol. 8, no. 4, pp. 192–202, 2023. doi: 10.54647/education880460

[20] L. Freina and M. Ott, "A literature review on immersive virtual reality in education: State of the art and perspectives," in *Proc. 11th Int. Conf. eLearning Softw. Educ.*, vol. 1, 2015, pp. 133–141. doi: 10.12753/2066-026x-15-020

[21] Y. J. Lin and T. J. Lan, "Language learning in virtual reality environments: Past, present, and future," *J. Educ. Technol. Soc.*, vol. 18, no. 4, pp. 486–497, 2015.

[22] X. Huang, D. Zou, G. Cheng, and H. Xie, "A systematic review of AR and VR enhanced language learning," *Sustain.*, vol. 13, no. 9, pp. 1–28, 2021. doi: 10.3390/su13094639

[23] G. W. Young, S. Stehle, B. Y. Walsh, and E. Tiri, "Exploring virtual reality in the higher education classroom: Using vr to build knowledge and understanding," *J. Univers. Comput. Sci.*, vol. 26, no. 8, pp. 904–928, 2020. doi: 10.3897/jucs.2020.049

[24] I. Nicolaidou, P. Pissas, and D. Boglou, "Comparing immersive virtual reality to mobile applications in foreign language learning in higher education: A quasi-experiment," *Interact. Learn. Environ.*, vol. 31, no. 4, pp. 2001–2015, 2023. doi: 10.1080/10494820.2020.1870504

[25] E. Korosidou, "The effects of augmented reality on very young learners' motivation and learning of the alphabet and vocabulary," *Digital*, vol. 4, no. 1, pp. 195–214, Feb. 2024. doi: 10.3390/digital4010010

[26] M. Dooly, T. Thrasher, and R. Sadler, "Whoa! Incredible!: Language learning experiences in virtual reality," *RELC J.*, vol. 54, no. 2, pp.

321–339, 2023. https://doi.org/10.1177/00336882231167610

[27] Y. Xie, T. Zhang, and H. Yang, "Effect of immersive virtual reality based upon input processing model for second language vocabulary retention," *Educ. Inf. Technol.*, pp. 1–21, 2025. https://doi.org/10.1007/s10639-025-13333-x

[28] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis, "Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis," *Comput. Educ.*, vol. 70, pp. 29–40, 2014. doi: 10.1016/j.compedu.2013.07.033

[29] W. R. Gall, M. D. Gall, and J. P. Borg, *Educational Research: An Introduction,* Boston.: Pearson Education Inc., 2003.

[30] H. Chia-Chien and B. A. Sandford, "The Delphi technique: Making sense of consensus," *Pract. Assessment, Res. Eval.*, vol. 12, no. 10, pp. 1–8, 2007. https://doi.org/10.7275/pdz9-th90

[31] A. H. Masnan, N. K. Ismail, W. Usop, and D. M. M. Zain, "Predicting preschool moral development with school climate: Fuzzy Delphi approach," *Int. J. Eval. Res. Educ.* , vol. 13, no. 3, pp. 1537–1544, 2024. doi: 10.11591/ijere.v13i3.27777

[32] P. A. Laksmiwati, Z. Lavicza, and A. N. Cahyono, "Empowering STEAM learning implementation through investigating Indonesian teacher experts' views with a Delphi method," *Indones. J. Learn. Adv. Educ.*, vol. 6, no. 2, pp. 214–229, 2024. doi: 10.23917/ijolae.v6i2.23460

[33] E. Toločka and A. Maceika, "Entrepreneurship and sustainability issues Delphi method application for evaluation of innovative product development scenario," vol. 11, no. 3, pp. 37–49, 2024.

[34] A. J. Candee. (2016). The classroom learning activities checklist: Validity evidence of an observation tool in preschool. [Online]. Available: https://www.proquest.com/openview/917148040aac25a796da3a5114fc3210/1?pq-origsite=gscholar&cbl=18750

[35] M. Appelbaum *et al.*, "Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report," *Am. Psychol.*, vol. 73, no. 1, pp. 3–25, 2018.

[36] C. Levitt, H. M. Bamberg, M. Creswell, J. W. Frost, D. M. Josselson, and R. Suárez-Orozco, "Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA publications and communications board task force report," *Am. Psychol.*, vol. 73, no. 1, pp. 26–46, 2018. https://doi.org/10.1037/amp0000151

[37] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Comput. Educ.*, vol. 147, 103778, 2020. doi: 10.1016/j.compedu.2019.103778

[38] O. Atabek, "Challenges in integrating technology into education," *Turkish Stud.—Inf. Technol. Appl. Sci.*, vol. 14, no. 1, pp. 1–19, 2019. doi: 10.7827/turkishstudies.14810

[39] H. Halimah, V. Nurviyani, A. Saepulah, J. Helmie, R. W. Wandawati, and S. Mutiah, "Virtual reality-supported drilling strategy in teaching English vocabulary to young learners," *Acitya J. Teach. Educ.*, vol. 4, no. 1, pp. 88–100, 2022. doi: 10.30650/ajte.v4i1.3195

[40] N. Kurniawati, A. Sofarini, E. H. Maolida, and R. T. D. Jatmika, "The praxis of integrating virtual reality into vocabulary teaching to young learners," *English Rev. J. English Educ.*, vol. 10, no. 2, pp. 371–380, 2022. doi: 10.25134/erjee.v10i2.6238

[41] M. Simonetti, D. Perri, N. Amato, and O. Gervasi, "Teaching math with the help of virtual reality," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12255 LNCS, pp. 799–809, 2020. doi: 10.1007/978-3-030-58820-5_57

[42] J. Parong and R. E. Mayer, "Learning science in immersive virtual reality," *J. Educ. Psychol.*, vol. 110, no. 6, pp. 785–797, 2018. doi: 10.1037/edu0000241

[43] J. Cohen, "A power primer," *Psychol. Bull.*, vol. 112, no. 1, pp. 155–159, 1992. doi: 10.1037/0033-2909.112.1.155

[44] M. R. Button, K. S. Ioannidis, J. P. A. Mokrysz, C. Nosek, B. A. Flint, J. Robinson, and E. S. J. Munafò, "Power failure: Why small sample size undermines the reliability of neuroscience," *Nat. Rev. Neurosci.*, vol. 14, no. 5, pp. 365–376, 2013. https://doi.org/10.1038/nrn3475