

Evaluating Student Learning Outcomes in Virtual Reality Adaptive Chemistry

Mohammad Farid Machfudin^{a,1*}, Esther Irawati Setiawan^{a,2}, Kevin Jonathan Halim^{a,3}, Joan Santoso^{a,4}, Agung Bella Putra Utama^{b,5}, Gunawan Gunawan^{a,6}, Samuel Budi Wardana Kusuma^{c,7}, Vrijraj Singh^{d,8}, Tong Nam Tuan Vu^{e,9}

^a Institut Sains dan Teknologi Terpadu Surabaya, Jl. Ngagel Jaya Tengah No.73-77, Surabaya, 60284 Indonesia

^b Universitas Negeri Malang, Jl. Semarang No.5, Sumberjaya, Kota Malang, 65145 Indonesia

^c Universitas Negeri Semarang, Jl. Raya Banaran, Kota Semarang, 50229 Indonesia

^d Agprop, Second floor, Passport seva kendra, Sub. Major Laxmi Chand Rd, Sector 18, Gurugram, Shahpur, Haryana 122015, India

^e AI Engineer, Change Interaction, Cao Thang Street, Ward Ban Co, Ho Chi Minh City, 70000 Vietnam

¹ mohammad.f22@mhs.istts.ac.id; ² esther@istts.ac.id; ³ kevin.j22@mhs.istts.ac.id; ⁴ joan@istts.ac.id;

⁵ agung.bella.2305349@students.um.ac.id; ⁶ gunawan@stts.edu; ⁷ samuelbudi@mail.unnes.ac.id; ⁸ vrijraj2396@gmail.com;

⁹ tongnamtuanvu@gmail.com

* corresponding author

ARTICLE INFO	ABSTRACT
Article history Received Oct 11, 2025 Revised Dec 15, 2025 Accepted Dec 19, 2025	This study evaluates the effectiveness of a Virtual Reality (VR)-based adaptive learning application in enhancing high school students' understanding of chemical compounds. The primary objective was to quantitatively assess the impact of the VR intervention on student learning outcomes across two distinct cohorts (N = 78). A pretest-posttest control-group design was employed, with two parallel groups (Group A and Group B) to ensure internal validity and comparability of results. The findings consistently indicate a marked contrast between the experimental and control conditions. Students in the control groups showed declines in performance, with negative learning gains of -8.32 and -15.20, suggesting learning loss when conventional instructional methods were used. In contrast, students exposed to the VR-based adaptive learning application demonstrated positive learning gains of +2.90 and +9.70, reflecting meaningful improvements in conceptual understanding. Further analysis of the intervention's impact revealed effect sizes ranging from medium (Cohen's $d = 0.722$) to very large ($d = 2.182$). These results indicate not only statistical significance but also substantial practical significance. Overall, the findings provide strong empirical evidence that the VR-based adaptive learning application is effective in preventing learning loss and significantly enhancing students' understanding of chemical compounds when compared to traditional instructional approaches. This is an open access article under the CC-BY license.
Keywords Adaptive Learning Compound Structure Virtual Reality Learning Outcome Evaluation Chemistry Education	



I. Introduction

Learning about chemical compounds is a significant challenge for high school students, mostly because molecular structures are abstract. This difficulty is supported by recent global education statistics. According to the OECD PISA 2022 results, about 66% of Indonesian students failed to reach the basic skill level (Level 2) in science. This is far behind the OECD average, where 76% of students reached this level. Furthermore, the report shows that almost no students in Indonesia were considered top performers (Level 5 or 6). This suggests there is a critical gap in students' ability to apply scientific knowledge creatively in new situations.

These difficulties are exacerbated by traditional teaching methods that rely heavily on static, two-dimensional (2D) diagrams in textbooks. Such methods often limit students' understanding of space and shape, forcing them to rely on imagination rather than seeing the models directly.

To solve these problems, Virtual Reality (VR) offers a powerful approach by creating immersive, interactive learning environments. Here, students can directly control and explore 3D molecular models. While previous research has shown that VR can increase engagement and motivation, most studies have focused on User Experience (UX) and interface usability rather than on actual learning outcomes.

Although high user satisfaction is an important first step, findings show that high engagement does not automatically lead to deep understanding or academic success. As a result, there is still uncertainty about whether VR is merely a motivational tool or a powerful support for learning. This research represents the next logical step in

this field: moving beyond checking usability to measuring actual learning improvements. There is a strong need to prove if VR can provide real educational benefits compared to traditional methods.

The main contribution of this article is the quantitative evaluation of a combined VR and adaptive learning system. Unlike standard VR applications, this proposed system uses adaptive algorithms to adjust the difficulty level for each student. This combination requires strong testing to prove its effectiveness. To carefully measure the impact, this study uses a pre-test and post-test control-group design to identify specific learning outcomes.

As a new method, this research also demonstrates an efficient way to create assessments by using Generative AI to produce chemistry test questions. This AI-driven approach ensures that high-quality questions can be produced easily and in large numbers. These questions were later reviewed by subject-matter experts to ensure they are suitable for educational use.

Therefore, the primary goal of this research is to investigate the effect of using modern technology in chemistry education. It examines whether a VR-based adaptive learning application can significantly improve students' chemistry learning outcomes compared with traditional teaching methods. Secondly, the study evaluates the validity of the AI-generated questions used to measure these achievements, ensuring that the research method is both innovative and scientifically correct.

Using Virtual Reality (VR) to learn science, especially chemistry, has been proven to help students understand difficult ideas (de Back et al., 2020; Song & Yu, 2024). By allowing students to interact with molecules in 3D space, VR can make complex topics such as stereochemistry and reaction mechanisms easier to understand. To build a high-quality VR learning application that leverages this benefit, we need a strong understanding of the key technologies and theories (Caneparo, 2001; Lui et al., 2023).

This application is built on a few key parts: VR to make the experience feel real, a game engine as the platform, and an adaptive learning method to personalize the learning experience for each student (Di Natale et al., 2020; Peirce et al., 2008). Adaptive learning systems are important because they adjust the pace and difficulty of the material based on how the student is performing (Liu, Kang, et al., 2017; Strielkowski et al., 2025). This can make a big difference in their success.

To assess the effectiveness of this application, we will use a pre-test and a post-test. This is a standard method in education to see how much knowledge a student has gained. The test questions will be generated using Generative AI, which can quickly create many types of questions (Clark, 2023; Mittal et al., 2024). However, this new method poses a major challenge: a human must carefully review all AI-generated questions to ensure they are accurate, fair, and appropriate for education.

A. Virtual Reality as an Immersive Learning Medium

Virtual Reality (VR) is a technology that allows users to enter and interact with a three-dimensional environment simulated by a computer (Burt & Louw, 2019). Using special devices such as a headset, users can experience the sensation of being physically present in the virtual environment (Bangay, Shaun, & Preston, Louise, 1998; Calabuig-Moreno et al., 2020). VR's ability to create immersive, interactive visual experiences makes it well-suited for educational applications (Alam & Ullah, 2016; Segura et al., 2020), especially for abstract concepts such as the structure of chemical compounds (Salzman et al., 1999; van Dam et al., 2002).

In conventional learning, students often struggle to visualize the shape of molecules from two-dimensional images in textbooks. VR overcomes this limitation by allowing students to see, rotate, and interact directly with 3D models of chemical compounds, thus providing a deeper and more intuitive spatial understanding (Hussain et al., 2024; Tan et al., 2022). This active and engaging learning experience has been shown to increase student interest and motivation (Setiawan et al., 2025).

The architecture of a VR system fundamentally consists of several key components:

- Input Processor: Processes position and orientation data from the headset's sensors and controllers.
- Simulation Processor: Runs the application's logic and processes user interactions within the virtual environment.
- Rendering Processor: Generates the visual and auditory output received by the user.
- World Database: Stores all objects and 3D assets present in the virtual world.

B. Unity Game Engine for Application Development

To realize an interactive and realistic VR environment, this research uses Unity as the primary game engine (Janecký et al., 2025; Lendalay et al., 2024). A game engine is a software framework that provides tools and functionality for developing interactive applications, ranging from graphics rendering and physics simulation to input management.

Unity was chosen for several key advantages:

- Flexibility and Multiplatform Support: Unity is known for its flexibility and supports development across platforms, including PC, consoles, mobile devices, and, most importantly, leading VR devices like the Meta Quest 2.
- Rapid and Efficient Development: With an intuitive interface, built-in assets, and an extensive asset store, Unity allows developers to create high-quality content more quickly and efficiently without having to build everything from scratch.

- Integrated XR Support: Unity provides a robust XR (Extended Reality) framework that simplifies development for VR, AR (Augmented Reality), and MR (Mixed Reality) applications. This enables seamless integration with VR hardware and the processing of data from the headset and controllers.

Development for the Meta Quest 2 headset is facilitated by enabling the Oculus provider in Unity's XR Plug-In Management menu, ensuring full compatibility and functionality.

C. Adaptive Learning for Personalized Instruction

Every student has a different learning pace and style. The "one-size-fits-all" approach to teaching is often ineffective. To address this, the application implements the adaptive learning method (Liu, McKelroy, et al., 2017; Wang et al., 2020).

Adaptive learning is an educational approach that uses technology to dynamically adjust the content, methods, and pace of learning to meet the unique needs of everyone (Dutta et al., 2024; Essa et al., 2023). The system collects data on a student's progress, strengths, and weaknesses, and then uses that information to:

- Adjust Difficulty Levels: The application automatically adjusts the material's difficulty based on the student's performance.
- Provide Personalized Recommendations: The system recommends the next level or material that is most suitable for the individual's learning needs.
- Increase Motivation: By presenting appropriate challenges that are neither too easy nor too hard, students tend to be more motivated and engaged in the learning process.

With this approach, each student can learn at their own pace, focus on areas they find difficult, and ultimately achieve more effective learning outcomes.

D. Cloud Firestore as a Supporting Database

To support adaptive learning, the application requires a reliable database system to store and manage user data. This research uses Cloud Firestore, a NoSQL database service from Google (Kesavan et al., 2023; Vera-Olivera et al., 2022).

Cloud Firestore was chosen for its ease of integration with Unity and its ability to synchronize data in real time across multiple devices. The stored data includes learning progress, quiz scores, interaction patterns, and user game history. This data serves as crucial input for the adaptive learning algorithm to function effectively, allowing the system to:

- Store placement test results to determine a student's initial level.

- Track student performance at each level to provide appropriate recommendations.
- Allow users to continue their learning sessions later with their progress saved.

E. Core Educational Content: Chemical Compounds

The learning content in this application focuses on the structure of chemical compounds, a fundamental topic in high school chemistry. The application covers three main types of chemical bonds:

- Ionic Bonds: Formed from the electrostatic attraction between positive (cations) and negative (anions) ions, typically between metal and non-metal atoms. An example is Sodium Chloride (NaCl).
- Covalent Bonds: Formed when two non-metal atoms share electron pairs to achieve a more stable configuration. An example is the water molecule (H₂O).
- Hydrocarbon Bonds: Compounds that consist exclusively of carbon (C) and hydrogen (H) elements. The simplest example is methane (CH₄).

By integrating these technologies, the proposed application aims to create a learning solution that is not only pedagogically effective but also engaging and motivating for students

II. Method

This research employed a quantitative experimental method to rigorously evaluate the effectiveness of the VR-based adaptive learning application. The study was structured to measure both the impact on student learning outcomes and the quality of the user experience. This section details the research design, participants, experimental procedure, and instruments used for data collection.

A. Research Design

The study used a pretest-posttest control-group design. This experimental framework is commonly used to evaluate the effect of an intervention by comparing a treatment group with a control group. In this context, the experimental group received the intervention: the VR chemistry learning application. The control group continued with their conventional learning methods. By comparing the score differences between the pre-test and post-test for both groups, the effectiveness of the VR application can be assessed quantitatively. This design allows researchers to better isolate the effect of the treatment, controlling for other external factors that might influence the results.

B. Participants and Procedure

The experiment included 78 tenth-grade students who tested the application. To ensure the evaluation was

reliable, the participants were separated into two main groups: Group A and Group B.

Group A consisted of 58 students, who were randomly assigned to two subgroups: an experimental group of 30 and a control group of 28. Meanwhile, Group B included 20 students who were divided equally, with 10 students in the experimental group and 10 students in the control group:

Pre-test Administration: Prior to the commencement of the instructional phase, a standardized pre-test was administered to both the experimental and control cohorts to establish a baseline of their chemical knowledge. This assessment consisted of 15 multiple-choice questions specifically designed to evaluate core competencies in chemical structures. The items covered a comprehensive range of topics, including identifying atomic numbers, deducing compound names from bond structures, distinguishing between bond types (Ionic vs. Covalent), and calculating the stoichiometry of Hydrocarbon chains. This initial measurement ensured that both groups started with a documented and comparable level of understanding before being subjected to different learning treatments.

The experimental group was guided to use the VR application for a controlled period of 15 minutes per student. This duration was strategically determined using Cognitive Load Theory in immersive environments. Previous research suggests that concise VR interventions (typically 10-20 minutes) are optimal for maximizing information retention while minimizing cognitive fatigue and cybersickness symptoms often associated with prolonged exposure (Makransky et al., 2019). Furthermore, given the high information density of 3D visualizations, students can grasp spatial concepts (Abdinejad et al., 2021)-such as molecular geometry-significantly faster than through traditional text-based instruction, rendering a 15-minute session sufficient for measurable learning gains. During this time, they experienced the core features of the application, including logging in, completing the tutorial level, viewing the cheat sheet, and playing through 2-3 game levels. Meanwhile, the control group followed its standard, conventional learning process for the same subject matter.

Post-test Administration: Upon completion of the respective learning sessions, both the experimental and control groups were immediately administered a post-test to measure learning outcomes. To ensure a valid comparative analysis of knowledge growth, this assessment was designed to be structurally equivalent to the pre-test, consisting of 15 multiple-choice questions covering the identical scope of chemical concepts, ranging from atomic valence identification and bond classification to the stoichiometry of hydrocarbon chains. This consistent testing framework enabled direct quantitative measurement of knowledge acquisition, allowing the precise calculation of the "learning gain" or "learning loss" experienced by students in each pedagogical environment.

C. Data Collection Instruments

Two primary types of instruments were used to collect data for this study: knowledge-based tests and a user experience questionnaire.

Pre-test and Post-test: These tests were designed to measure the students' understanding of chemistry concepts covered in the application. Each test consisted of 15 questions that covered topics such as identifying atomic numbers, determining the type of bond in a compound (ionic, covalent, or hydrocarbon), and understanding how atoms combine to form compounds. The pre-test measured initial knowledge, while the post-test measured the knowledge gained after the intervention, allowing for a direct comparison of learning outcomes between the two groups.

D. Game Flow and User Interaction Design

The application workflow is designed to create a personalized learning path for every student. The process begins with the Initialization Phase. When students launch the VR application, they log in using a unique username. If the system identifies a new user, it automatically activates a Placement Test. This diagnostic test contains 10 questions covering general chemistry knowledge, as well as specific topics like ionic, covalent, and hydrocarbon bonds.

The test results are used by an adaptive algorithm to measure the student's current skill level. Based on this data, the system recommends the most suitable difficulty level-whether Easy, Normal, or Hard. This step ensures the student does not feel that the lesson is too difficult or too easy right from the start. After this initial assessment is finished, the student proceeds to the Core Gameplay Loop, which serves as the main learning activity. This phase takes place within the 'Play Level' mode and follows a structured interactive sequence.

Task Generation: The system generates a set of chemical compounds as "quests" for the player to build. These questions are randomized but filtered by the adaptive engine to match the student's current mastery level.

Exploration and Collection: Students navigate the 3D virtual laboratory using VR controllers. They must locate specific atom capsules scattered throughout the environment and extract the necessary atoms (e.g., Hydrogen, Oxygen, Carbon) by physically grabbing them with the controller's grip button.

Construction and Assembly: Students bring the collected atoms to the "Combining Area". Here, they construct molecules by physically bringing atoms close together to form bonds. The interface allows users to toggle between adding bonds (single, double, or triple) and removing them, visualizing the formation of Ionic or Covalent bonds in real time.

Validation: The system continuously monitors the structure. Once the student believes the molecule is complete, the system validates the arrangement of atoms and bond types against the correct chemical formula. If correct, the task is marked as complete.

The final stage of the process is the Evaluation and Adaptive Feedback Loop. After the student successfully builds the required chemical compounds, the session ends with a 'Bonus Round.' In this reverse challenge, the system shows a complete 3D molecule, and the student must identify its name from multiple-choice options. This activity helps to strengthen their recognition skills.

When the level is finished, the 'Game Over' screen displays detailed performance data, including the completion time, accuracy score, and the number of hints used. This part is important because it completes the adaptive learning cycle: the algorithm compares these results against the level's requirements. If the student meets the target score, the system recommends moving to a higher difficulty or a new topic (Vertical Transition). If the score is not high enough, the system suggests a review level or repeating the current topic (Horizontal Transition) to ensure the student understands the concepts before moving forward.

E. Adaptive Learning Mechanism & Transition Model

The application's main logic is based on an Adaptive Course Delivery model, which adjusts educational content to each student's skill level in real time. To determine the best starting point, the system uses a specific scoring method during the initial placement test. This test contains 10 questions: 1 on general chemistry and 3 specific questions for each bond type (Ionic, Covalent, and Hydrocarbon).

The scoring system distinguishes between general and specific knowledge. General questions give 10 points to all categories, while specific questions give 30 points only to the related category. The system then compares these scores against a target of 70 points. If a student scores below 70 in the first category, they start at the Ionic Level. However, if they score above this target, the algorithm checks the next category. This allows skilled students to skip basic content, preventing them from getting bored by unnecessary repetition.

A key part of the progression system is the Dual-Transition Model, which controls how a student moves through the lessons based on their performance. The first path, Vertical Transition, is used for advancement. It activates when a student proves they understand the current topic by reaching the target score on 'Hard' difficulty. When successful, the system unlocks the next concept, such as moving from Ionic to Covalent compounds.

On the other hand, the system uses a Horizontal Transition for review and practice. If a student does not

reach the target score, they cannot move up to the next level. This prevents the material from becoming too overwhelming. In this situation, the student stays on the same topic, but the system adjusts the difficulty or mixes the questions to help them better understand without introducing new, complex ideas.

To ensure that these difficulty thresholds are objective and scalable rather than arbitrary, the system calculates target scores using a statistical Quartile Formula (Q_n). The target score for a level is derived by multiplying the total number of chemical bonds (M) required to complete the tasks by the base score per bond (100). The formula is defined as

$$Q_n = \frac{n(M+1)}{4} \times 100$$

where n represents the difficulty tier 5. Specifically, "Easy" mode sets the target at the 1st Quartile ($n=1$), "Normal" mode at the Median ($n=2$), and "Hard" mode at the 3rd Quartile ($n=3$). This mathematical approach ensures that "Hard" mode requires significantly higher accuracy and completion rates than "Easy" mode, providing a standardized metric for mastery across different chemical topics.

Beyond general difficulty adjustment, the system's adaptability extends to specific content remediation through a "Todo List" Mechanism. When a student struggles with specific compounds, for example, failing to correctly construct $FeCl_3$, the system flags these items in the database. During the subsequent Horizontal Transition, the random question generation algorithm prioritizes these flagged items, ensuring they reappear in the problem set 6. This ensures that students are explicitly retested on their weak points rather than receiving a completely random set of easy questions, thereby facilitating targeted remediation and ensuring comprehensive understanding before progression.

F. Chemical Structure Validation Engine

To maintain high educational standards, the application includes a strong Structure Validation Engine that acts as a real-time checker. Instead of simply matching visuals, this engine uses a graph-based algorithm to compare the molecule the user builds against the correct chemical formula. This checking process runs a detailed review every time the user adds or changes a bond.

The first step is the Quantity Check. Here, the system compares the total number of atoms and bonds in the user's workspace with the target formula. This initial step prevents wasted processing time. If the user has the wrong number of atoms, for example, using three Hydrogen atoms for a water molecule instead of two, the system immediately marks the structure as incorrect without doing further analysis.

Once the quantities match, the algorithm moves to the Bond Structure Verification stage. In this phase, the

system checks every bond the user creates to confirm connections and bond types. It verifies if the right atoms are connected (for example, connecting Carbon to Hydrogen) and checks the Bond Type (Single, Double, or Triple). This step is important for distinguishing between compounds that have similar atoms but different structures, such as ensuring double bonds are in the right place for an Ethene molecule.

Finally, the engine runs a Composition Consistency Check. Even if the structure looks correct, the system performs a final review to ensure the types of atoms match the answer key exactly. Using a counting method, the algorithm verifies that the number of each specific element (e.g., exactly 1 Sodium and 1 Chlorine for NaCl) matches the target formula exactly. Only when the structure passes all these checks-quantity, structure, and composition-does the system activate the 'is Correct' signal, marking the task as complete and unlocking the next part of the level.

G. Dynamic Scoring Algorithm for Cognitive Efficiency

To evaluate student performance beyond binary "correct or incorrect" outcomes, the application employs a Dynamic Scoring Algorithm designed to measure cognitive efficiency and discourage trial-and-error behavior. As detailed in the procedural design, the final score calculation is a composite derivative of multiple gameplay variables, not merely the completion of tasks. The algorithm begins by calculating a Base Score, in which every correct chemical bond formed within a target compound awards the student 100 points, directly incentivizing mastery of complex molecular structures.

However, to ensure the score reflects true proficiency, the system applies specific deductions for inefficiency. A Hint Penalty subtracts 30 points for each instance the student relies on the assistance system, thereby rewarding independent problem-solving. Furthermore, to discourage the random collection of assets, an Excess Atom Penalty is enforced; the system calculates the difference between the number of atoms collected (*atomsCollected*) and the actual number required (*atomsNeeded*), deducting 10 points for every unnecessary atom held by the player. Conversely, speed and deep understanding are rewarded: the remaining time (*timeLeft*) is converted into points with a multiplier of 5, and successfully answering the post-level "Bonus Question" grants a flat addition of 100 points. This mathematical framework ensures that the adaptive engine receives a high-fidelity metric of student capability, distinguishing between a student who struggles through trial-and-error and one who solves problems with precision and speed.

III. Results and Discussion

This chapter presents a detailed statistical analysis of the data obtained from the pretest-posttest experimental design. To ensure the robustness of the findings, data were

collected from two distinct groups: Group A (N=58) and Group B (N=20). The objective is to determine the effectiveness of the VR-based adaptive learning application by analyzing student learning outcomes across these cohorts.

A. Descriptive Statistics

The descriptive statistics provide a comprehensive overview of student performance across two distinct experimental cohorts: Group A and Group B. The data highlights a divergent trajectory between the experimental and control groups in both settings, offering initial insight into the intervention's impact.

To facilitate a clear interpretation of these findings, the tables present several key statistical metrics. The "N" column denotes the sample size, ensuring that the comparison groups are balanced. The "Mean" serves as the primary indicator of the central tendency of student scores, reflecting average performance before and after the intervention. The "Max" and "Min" columns outline the score range, highlighting the highest and lowest achievements to show how the intervention affected students across the entire ability spectrum. Finally, the "Std. Dev." (Standard Deviation) measures the variability of the scores; a lower value indicates that student performance is clustered closely around the average, while a higher value suggests a wider disparity in student understanding.

Table 1 details the results for Group A (N=58). The experimental group (N=30) demonstrated a positive learning trajectory, with the mean score rising from 57.47 (SD=13.98) in the pretest to 60.37 (SD=14.60) in the posttest. This indicates a steady acquisition of knowledge using the VR application.

In contrast, the control group (N=28) exhibited a concerning regression. Although they started with a higher baseline knowledge (Mean=68.39), their performance significantly deteriorated during the conventional learning process, dropping to a posttest mean of 60.07. This decline suggests that without the visualization support provided by VR, students struggled to retain or correctly apply abstract chemical concepts over time.

Table 1. Descriptive Statistics for Experimental and Control Groups (Group A)

Variable	N	Mean	Max	Min	Std. Dev.
Experiment Pretest	30	57.47	87	33	13.98
Experiment Posttest	30	60.37	87	33	14.60
Control Pretest	28	68.39	93	33	13.79
Control Posttest	28	60.07	87	27	18.20

Table 2 presents the data for Group B (N=20), which corroborates the trends observed in Group A with even greater intensity. The experimental group (N=10)

achieved a substantial improvement, increasing their mean score from 42.20 to 51.90.

Conversely, the control group (N=10) experienced a sharp decline in performance. Their mean score plummeted from 32.60 in the pretest to 17.40 in the posttest. This drastic drop reinforces the observation that traditional methods may be insufficient for maintaining student understanding in this specific topic, whereas the VR intervention successfully facilitated knowledge growth.

Table 2. Descriptive Statistics for Experimental and Control Groups (Group B)

Variable	N	Mean	Max	Min	Std. Dev.
Experiment Pretest	10	42.20	67	20	16.00
Experiment Posttest	10	51.90	73	27	16.35
Control Pretest	10	32.60	53	13	11.48
Control Posttest	10	17.40	33	7	9.45

B. Prerequisite Test Results

Before proceeding with the hypothesis testing using the Independent Samples T-test, it is imperative to verify that the data satisfy the underlying assumptions of parametric statistics. Two critical assumptions were tested: the normality of the data distribution and the homogeneity of variances between the groups. The Shapiro-Wilk test was employed to assess whether the pretest and posttest scores for both the experimental and control groups were normally distributed (Monter-Pozos & Gonz^{lez}-Estrada, 2024), with the criterion for normality being a significance value (p-value) greater than 0.05.

For Group A, as detailed in Table 3, the analysis confirmed that the data were normally distributed. Specifically, the experimental group yielded p-values of 0.341 for the pretest and 0.232 for the posttest, while the control group yielded 0.350 and 0.175, respectively. Since all obtained values exceeded the 0.05 significance threshold, the assumption of normality was satisfied.

Table 3. Normality Test Results (Group A)

Group	Pretest (p-value)	Posttest (p-value)
Experimental	0.341	0.232
Control	0.350	0.175

A similar pattern was observed in Group B, as presented in Table 4. The experimental group demonstrated p-values of 0.299 and 0.248, whereas the control group recorded values of 0.884 and 0.225. These results indicate no significant deviation from a normal distribution across any of the subgroups in the second cohort.

Table 4. Normality Test Results (Group B)

Group	Pretest (p-value)	Posttest (p-value)
Experimental	0.299	0.248
Control	0.884	0.225

Following the normality check, Levene's Test for Equality of Variances was conducted to determine if the experimental and control groups had equal variances. This assumption is considered valid if the p-value is greater than 0.05. For Group A, the analysis yielded Levene's test statistics of 0.702 for the pretest and 0.168 for the posttest, as shown in Table 5. Both values are well above 0.05, confirming that the variance between the two groups is homogeneous. Similarly, for Group B shown in Table 6, the test yielded p-values of 0.140 for the pretest and 0.220 for the posttest. Thus, the assumption of homogeneity of variance is also fulfilled for the second cohort. Consequently, since the data from both Group A and Group B successfully met the assumptions of normality and homogeneity, the use of the Independent Samples T-test to compare the learning outcomes is statistically valid and appropriate.

Table 5. Homogeneity of Variance Test Results (Group A)

Variable	Levene Statistic	p-value
Pretest	0.148	0.702
Posttest	1.951	0.168

Table 6. Homogeneity of Variance Test Results (Group B)

Variable	Levene Statistic	p-value
Pretest	2.387	0.140
Posttest	1.617	0.220

C. Hypothesis Testing Results

To empirically verify the intervention's effectiveness, hypothesis testing was conducted using Independent-Samples T-Tests. The primary variable analyzed was the "gain score," which represents the difference between the posttest and pretest scores for each participant (Gain = Posttest - Pretest). This metric isolates the actual learning progress (or regression) attributable to the intervention period, effectively controlling for initial differences in baseline knowledge.

The results for the first cohort are presented in Table 7. The data reveals a striking divergence in learning outcomes between the two conditions. The experimental group, which utilized the VR application, achieved a positive mean gain of +2.90 (SD=14.91), indicating a net improvement in understanding. Conversely, the control group exhibited a significant "learning loss," recording a negative mean gain of -8.32 (SD=16.20).

The Independent Samples T-test yielded a t-value of 2.747 and a significance value (p-value) of 0.008. Since the obtained p-value is well below the standard alpha level

of 0.05 ($p < 0.05$), the null hypothesis is rejected. This confirms that the difference in learning outcomes between the VR users and the conventional learners in Group A is statistically significant, favoring the VR intervention.

Independent Samples T-test Results on Gain Scores (Group A)

Group	N	Mean Gain	Std. Dev.	t-value	p-value (2-tailed)
Experimental	30	+2.90	14.91	2.747	0.008
Control	28	-8.32	16.20		

The findings from the second cohort, detailed in Table 8, corroborate and amplify the trends observed in Group A. In this group, the experimental participants demonstrated a substantial improvement with a mean gain of +9.70 (SD=11.02). In sharp contrast, the control participants experienced a severe decline in performance, recording a mean gain of -15.20 (SD=11.79).

The statistical analysis for Group B yielded a higher t-value of 4.879 and a p-value of 0.000. This result provides overwhelmingly strong evidence against the null hypothesis. The practically zero p-value signifies that the probability of observing such a large difference in performance by random chance is negligible. Therefore, the positive impact of the VR intervention in Group B is deemed highly significant.

Table 7. Independent Samples T-test Results on Gain Scores (Group B)

Group	N	Mean Gain	Std. Dev.	t-value	p-value (2-tailed)
Experimental	10	+9.70	11.02	4.879	0.000
Control	10	-15.20	11.79		

To quantify the magnitude of the intervention's impact, Cohen's d was calculated for both cohorts, revealing a substantial educational benefit. For Group A, the analysis yielded a Cohen's d of 0.722, which indicates a medium-to-large effect size. In Group B, the impact was exceptional, resulting in a Cohen's d of 2.182, classified as a very large effect. This suggests that the specific combination of immersive VR with adaptive learning algorithms used in this study offers greater efficacy than standard VR implementations.

The results from these two distinct datasets provide strong, corroborating evidence of the application's effectiveness. A striking and consistent finding across both groups is the negative gain observed in the control groups, with values of -8.32 for Group A and -15.20 for Group B. This "learning loss" phenomenon is critically important; it indicates that without appropriate visualization tools, students struggling with the abstract nature of chemical compounds may experience a degradation in understanding or an increase in misconceptions over time when relying solely on conventional 2D methods.

In contrast, the VR intervention successfully reversed this trend across both trials. In the larger Group A, the application stabilized learning and produced a meaningful positive effect ($d=0.722$), while in Group B, it demonstrated a profound impact ($d=2.182$) by significantly boosting scores, while the control group's performance plummeted. This dual validation underscores the technology's capability to act as a crucial cognitive scaffold. By matching the complexity of 3D models to the students' proficiency level, the adaptive system effectively prevented the cognitive overload often associated with complex spatial topics, facilitating measurable learning gains where traditional methods failed to sustain student understanding.

D. Field Implementation and Observational Findings

To qualitatively validate the intervention process, direct observations were conducted during the experimental sessions held at the school's audiovisual laboratory. The implementation followed a rigorous 15-minute protocol designed to systematically transition students from passive observers to active participants. To ensure safety and facilitate real-time classroom supervision, the experimental setup used a seated VR configuration.

The session commenced with the Initialization Phase. As shown in Fig. 1, the VR headset's visual output was mirrored to a laptop screen via casting. This setup was critical to the research methodology, as it allowed the instructor to monitor students' progress in real time without interrupting their immersion. From this dashboard, the instructor verified that the Adaptive Algorithm correctly triggered the appropriate difficulty level based on the student's initial placement test results.

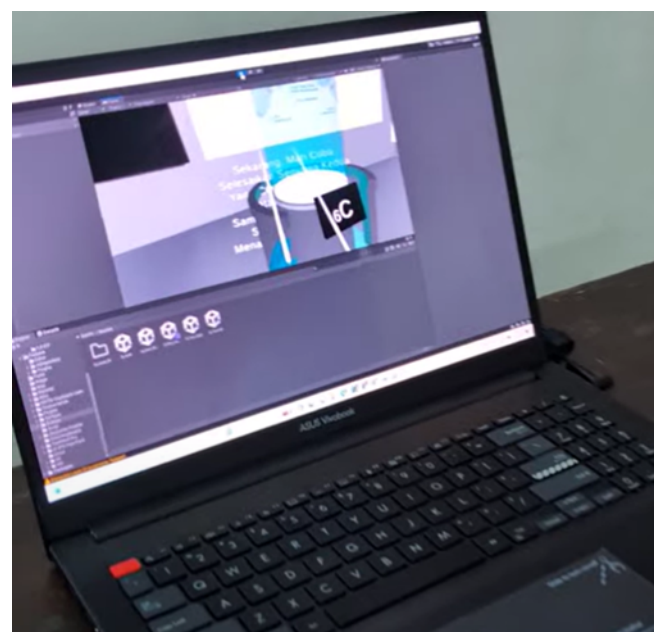


Fig. 1. Sustained Engagement and Focus.

Following initialization, students engaged with the Tutorial Module. Fig. 2 illustrates the seated interaction model. Although stationary, the learning process remained highly active. The student is depicted utilizing the handheld controllers to navigate the virtual interface. Observations at this stage confirmed that students quickly adapted to the input mechanics, specifically using the "Grip" and "Trigger" buttons to grasp virtual atoms, demonstrating that the seated configuration did not hinder the application's kinesthetic learning.



Fig. 2. Initial Setup and Instructor Supervision.

During the core Adaptive Gameplay Phase, behavioral engagement peaked. Fig. 3 captures the high level of focus and immersion exhibited by participants. In this phase, students were tasked with constructing complex molecular structures (e.g., Hydrocarbons) as dictated by the adaptive engine. Unlike passive traditional learning, this observation confirms that the VR application successfully facilitated sustained attention and active cognitive processing, even within a seated safety protocol.



Fig. 3. Sustained Engagement and Focus.

Collectively, these observational findings provide crucial context for the statistical results presented later in this study. The seamless transition from the initial tutorial to the complex adaptive gameplay confirmed the high usability of the interface, even for first-time VR users. Furthermore, the visible enthusiasm and sustained attention observed across all participants validate the hypothesis that immersive technology can effectively bridge the gap between abstract theory and concrete understanding. The absence of significant technical hurdles or cybersickness during the seated sessions suggests that the 15-minute intervention protocol is not only pedagogically effective but also practically viable for integration into standard classroom environments. This qualitative evidence strongly supports the conclusion that the reported learning gains were driven by a heightened state of cognitive engagement that traditional methods struggle to replicate."

IV. Conclusion

The comprehensive evaluation detailed in this paper, conducted across two distinct cohorts (Group A and Group B), provides robust empirical evidence of the VR-based adaptive learning application's efficacy in teaching chemical compound structures. The results reveal a decisive divergence in learning trajectories between the teaching methods. The statistical analysis confirmed that the differences in learning outcomes were significant ($p=0.008$ for Group A and $p=0.000$ for Group B), favoring the VR intervention in both cases. While the experimental groups consistently achieved positive mean gains in knowledge acquisition of +2.90 and +9.70, respectively, the control groups in both cohorts showed notable regression, with negative mean gains of -8.32 and -15.20. This "learning loss" phenomenon offers a critical insight: reliance on traditional, static 2D instructional methods for abstract spatial topics may not only fail to improve understanding but can lead to the degradation of students' initial knowledge due to misconceptions. The VR intervention successfully mitigated this issue by acting as a cognitive anchor. The large effect sizes observed ($d=0.722$ and $d=2.182$) further quantify the practical magnitude of this benefit, confirming that immersive visualization combined with adaptive scaffolding significantly outperforms conventional instruction in sustaining and improving student understanding. Furthermore, the study successfully validated a novel methodological workflow by integrating Generative AI for assessment creation. This approach proved highly efficient, producing test items that, after expert validation, demonstrated strong psychometric properties. The success of this research highlights the powerful synergy between immersive educational technologies, adaptive algorithms, and AI-driven content development. Collectively, these findings suggest that such integrated systems have the potential to fundamentally transform science education,

shifting the paradigm from passive information retention to active, sustained conceptual mastery.

References

- Abdinejad, M., Talaie, B., Qorbani, H. S., & Dalili, S. (2021). Student Perceptions Using Augmented Reality and 3D Visualization Technologies in Chemistry Education. *Journal of Science Education and Technology*, 30(1), 87–96. <https://doi.org/10.1007/s10956-020-09880-2>
- Alam, A., & Ullah, S. (2016). Adaptive 3D-Virtual Learning Environments: From Students' Learning Perspective. *2016 International Conference on Frontiers of Information Technology (FIT)*, 7–10. <https://doi.org/10.1109/FIT.2016.010>
- Bangay Shaun, & Preston Louise. (1998). *AN investigation into factors influencing immersion in interactive virtual reality environments*. <https://doi.org/10.3233/978-1-60750-902-8-43>
- Burt, M., & Louw, C. (2019). Virtual reality enhanced roller coasters and the future of entertainment – audience expectations. *World Leisure Journal*, 61(3), 183–199. <https://doi.org/10.1080/16078055.2019.1639274>
- Calabuig-Moreno, F., González-Serrano, M. H., Fombona, J., & García-Tascón, M. (2020). The Emergence of Technology in Physical Education: A General Bibliometric Analysis with a Focus on Virtual and Augmented Reality. *Sustainability*, 12(7), 2728. <https://doi.org/10.3390/su12072728>
- Caneparo, L. (2001). Shared virtual reality for design and management: the Porta Susa project. *Automation in Construction*, 10(2), 217–228. [https://doi.org/10.1016/S0926-5805\(99\)00032-1](https://doi.org/10.1016/S0926-5805(99)00032-1)
- Clark, T. M. (2023). Investigating the Use of an Artificial Intelligence Chatbot with General Chemistry Exam Questions. *Journal of Chemical Education*, 100(5), 1905–1916. <https://doi.org/10.1021/acs.jchemed.3c0-0027>
- de Back, T. T., Tinga, A. M., Nguyen, P., & Louwerse, M. M. (2020). Benefits of immersive collaborative learning in CAVE-based virtual reality. *International Journal of Educational Technology in Higher Education*, 17(1), 51. <https://doi.org/10.1186/s41239-020-00228-9>
- Di Natale, A. F., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology*, 51(6), 2006–2033. <https://doi.org/10.1111/bjet.13030>
- Dutta, S., Ranjan, S., Mishra, S., Sharma, V., Hewage, P., & Iwendi, C. (2024). Enhancing Educational Adaptability: A Review and Analysis of AI-Driven Adaptive Learning Platforms. *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, 1–5. <https://doi.org/10.1109/ICIPTM59628.2024.10563448>
- Essa, S. G., Celik, T., & Human-Hendricks, N. E. (2023). Personalized Adaptive Learning Technologies Based on Machine Learning Techniques to Identify Learning Styles: A Systematic Literature Review. *IEEE Access*, 11, 48392–48409. <https://doi.org/10.1109/ACCESS.2023.3276439>
- Hussain, R., Sabir, A., Lee, D.-Y., Zaidi, S. F. A., Pedro, A., Abbas, M. S., & Park, C. (2024). Conversational AI-based VR system to improve construction safety training of migrant workers. *Automation in Construction*, 160, 105315. <https://doi.org/10.1016/j.autcon.2024.105315>
- Janecký, D., Kučera, E., Haffner, O., & Hricková, L. (2025). Development of an Interactive VR Application Using Unity: Its Impact on Learning and Performance. *2025 Cybernetics & Informatics (K&I)*, 1–6. <https://doi.org/10.1109/KI64036.2025.10916426>
- Kesavan, R., Gay, D., Thevessen, D., Shah, J., & Mohan, C. (2023). Firestore: The NoSQL Serverless Database for the Application Developer. *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 3376–3388. <https://doi.org/10.1109/ICDE-55515.2023.00259>
- Lendalay, A. R., Pabba, R., Katta, S. R., Ahmed, M. A., Karthik, R. U., & Prasad, K. S. (2024). Development and Implementation of a Virtual Reality Simulation for an Immersive College Environment Using the Unity Game Engine. *2024 IEEE North Karnataka Subsection Flagship International Conference (NKCon)*, 1–8. <https://doi.org/10.1109/NKCon-62728.2024.10774914>
- Liu, M., Kang, J., Zou, W., Lee, H., Pan, Z., & Corliss, S. (2017). Using Data to Understand How to Better Design Adaptive Learning. *Technology, Knowledge and Learning*, 22(3), 271–298. <https://doi.org/10.1007/s10758-017-9326-z>
- Liu, M., McKelroy, E., Corliss, S. B., & Carrigan, J. (2017). Investigating the effect of an adaptive learning intervention on students' learning. *Educational Technology Research and Development*, 65(6), 1605–1625. <https://doi.org/10.1007/s11423-017-9542-1>
- Lui, A. L. C., Not, C., & Wong, G. K. W. (2023). Theory-Based Learning Design with Immersive Virtual Reality in Science Education: a Systematic Review. *Journal of Science Education and Technology*, 32(3), 390–432. <https://doi.org/10.1007/s10956-023-10035-2>
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation increases presence but reduces learning. *Learning and Instruction*, 60, 225–236. <https://doi.org/10.1016/j.learninstruc.2017.12.007>
- Mittal, U., Sai, S., Chamola, V., & Sangwan, D. (2024). A Comprehensive Review on Generative AI for Education. *IEEE Access*, 12, 142733–142759. <https://doi.org/10.1109/ACCESS.2024.3468368>
- Monter-Pozos, A., & González-Estrada, E. (2024). Testing the skew normal distribution using the Shapiro–Wilk test. *Journal of Computational and Applied Mathematics*, 440, 115649. <https://doi.org/10.1016/j.cam.2023.115649>
- Peirce, N., Conlan, O., & Wade, V. (2008). Adaptive Educational Games: Providing Non-invasive Personalised Learning Experiences. *2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning*, 28–35. <https://doi.org/10.1109/DIGITEL.2008.30>

- Salzman, M. C., Dede, C., Loftin, R. B., & Chen, J. (1999). A Model for Understanding How Virtual Reality Aids Complex Conceptual Learning. *Presence: Teleoperators and Virtual Environments*, 8(3), 293–316. <https://doi.org/10.1162/105474699566242>
- Segura, R. J., del Pino, F. J., Ogáyar, C. J., & Rueda, A. J. (2020). VR-OCKS: A virtual reality game for learning the basic concepts of programming. *Computer Applications in Engineering Education*, 28(1), 31–41. <https://doi.org/10.1002/cae.22172>
- Setiawan, E. I., Machfudin, M. F., Saputra, D. G., Santoso, J., Gunawan, G., & Kusuma, S. B. W. (2025). Evaluating User Experience of a Virtual Reality-Based Adaptive Learning Application on Chemical Compound Structures for High School Students. *International Journal of Engineering, Science and Information Technology*, 5(4), 427–433. <https://doi.org/10.52088/ijesty.v5i4.1445>
- Song, H. W., & Yu, J. (2024). Effects of Augmented Reality-based Dual-Task Proprioceptive Training on Postural Stability, Positioning Sensation, and Cognition. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 15(1), 88–103. <https://doi.org/10.58346/JOWUA.2024.II.007>
- Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., & Vasileva, T. (2025). AI-driven adaptive learning for sustainable educational transformation. *Sustainable Development*, 33(2), 1921–1947. <https://doi.org/10.1002/sd.3221>
- Tan, Y., Xu, W., Li, S., & Chen, K. (2022). Augmented and Virtual Reality (AR/VR) for Education and Training in the AEC Industry: A Systematic Review of Research and Applications. *Buildings*, 12(10), 1529. <https://doi.org/10.3390/buildings12101529>
- van Dam, A., Laidlaw, D. H., & Simpson, R. M. (2002). Experiments in Immersive Virtual Reality for Scientific Visualization. *Computers & Graphics*, 26(4), 535–555. [https://doi.org/10.1016/S0097-8493-\(02\)00113-9](https://doi.org/10.1016/S0097-8493-(02)00113-9)
- Vera-Olivera, H., Guo, R., Huacarpuma, R. C., Da Silva, A. P. B., Mariano, A. M., & Holanda, M. (2022). Data Modeling and NoSQL Databases - A Systematic Mapping Review. *ACM Computing Surveys*, 54(6), 1–26. <https://doi.org/10.1145/3457608>
- Wang, S., Christensen, C., McBride, E., Kelly, H., Cui, W., Tong, R., Shear, L., Yarnall, L., & Feng, M. (2020). Identifying Gaps in Use of and Research on Adaptive Learning Systems. *Proceedings of the 12th International Conference on Computer Supported Education*, 118–124. <https://doi.org/10.5220/000959-0701180124>