Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

ISSN ONLINE: 2447-0228

Manaus, v.12 n.58, p.1058-1071. March/April, 2026.
DOI: https://doi.org/10.5935/jetia.v12i58.3257

RESEARCH ARTICLE OPEN ACCESS

HCAVR-PLE: A DUAL-PHASE ADAPTIVE AI-VR FRAMEWORK FOR EQUITABLE, EXPLAINABLE, AND ENGAGING PHYSICS EDUCATION

Premalatha. R[1], Indu. H*[2]

[1]Researcher, Department of Education , Avinashilingam Institute for Home Science and Higher Education for Women , Tamil Nadu, India.
[2]Professor and Dean, School of Education, Avinashilingam Institute for Home Science and Higher Education for Women, Tamil Nadu, India

[1]https://orcid.org/0000-0002-9688-1257, [2]http://orcid.org/0000-0003-0562-101X

Email: 21phedf001@avinuty.ac.in, indu_edu@avinuty.ac.in

ARTICLE INFO

Article History
Received: January 8, 2026
Reviewed: February 8, 2026
Accepted: March 10, 2026
Published: April 30, 2026

Keywords:
Adaptive Learning,
Human-Centered AI,
Virtual Reality in Education,
Equity in STEM,
Secondary Physics Education.

ABSTRACT

There is still an issue of equity and involvement in secondary STEM education, especially among diverse students dealing with the concepts of physics in resource-limited environments. Traditional teaching methods, such as traditional virtual reality (VR), tend to be non-personalized, restrictive in inclusivity, and thus ineffectual. The study proposes HCAVR-PLE, a Human-Centered Adaptive Virtual Reality Personalized Learning Environment that combines the Cognitive Affective Model of Immersive Learning (CAMIL), multi-armed bandit algorithms, explainable AI (XAI), and reflective prompts to provide clear, bias-free teaching. HCAVR-PLE had enormous effects, where the AI-VR (Real) had an increase in PAT of 29.2% compared to the non-adaptive groups (η^2=0.39, p<0.05). CAMIL engagement was greatest in adaptive AI-VR (η^2=0.52) and gender/SES equity gaps were least in the AI-VR (Real) condition. The framework was validated using the dual-phase method: 300 synthetic learner profiles (Phase I) based on PISA and OULAD data sets, and 221 Grade 8 students (Phase II) in the state of Tamil Nadu, India and compared to using Teacher dashboards, built based on HG-SCM-based explainable AI, made delivery of personalized experiences interpretable and equitable, whereas Phase II had established efficacy of framework without intervention of dashboard. These results reveal the potential of HCAVR-PLE to revolutionise STEM education by delivering greater engagement and equity solutions, potentially providing a scalable, universal classroom model of learning.

I. INTRODUCTION

Equity (and engagement) are an ongoing issue in science teaching, and decreasing motivation is of intersection with abstract conceptual learning in middle school physics [1]. Although VR enriches traditional instruction in some yet partial areas, it may be less efficacious (or not efficacious at all) over time, at least on less adaptive and inclusive fronts (than VR) due to the variability of gendered and SES-diverse learners [2-5]. Despite the potential of VR to increase immersion and enjoyment, the static implementation will not react to the changing cognitive and affective states of learners [5]. Recent literature cites a serious lack of empirical research on adaptive, explainable AI as it relates to STEM education in K-12 settings, with the existing research being limited to conceptual models [6-8] and collegiate settings [9],[10]. Furthermore, the number of frameworks, which would blend affective engagement models, such as CAMIL [11] and transparency mechanisms, still remains negligible [12],[13]. The new adaptive systems with a human focus on AI promise to be a good alternative. Combining engagement detection (e.g., CAMIL) and Explainable AI (XAI) helps create more transparent, fair, and learner-responsive settings [14],[7],[15]. The framework uses major theories of education to enhance its pedagogical basis. The concept of adaptive sequencing is based on the Zone of Proximal Development by Vygotsky, since the content being taught is aligned to the state of readiness of the learners.

Affective tagging that is provided by CAMIL leads to the minimization of extraneous cognitive load, which aligns with the ideas of Cognitive Load Theory. Individualization and reflective cues promote feelings of autonomy, competence, and relatedness, which are in line with the Self-Determination Theory. All these connections combine to make HCAVR-PLE a theory-informed mode of education in addition to a technological innovation. The literature is growing in demand for intelligent systems that put learner agency, ethical AI use, and explainable personalization pathways, in addition to the need to ameliorate performance [16],[12],[13]. The HCAVR-PLE is a two-phase AI-VR-based learning framework that is specific to secondary-level physics. These aims are to simulate learner performance and engagement patterns using synthetic data grounded on PISA and OULAD [1],[17], to deploy an AI-VR system in which the instruction can be adaptive to affective cues as modeled on the CAMIL, and to evaluate empirically how this framework can affect achievement, equity, engagement reflective thinking, and explainability on both simulated learners and live learners. This study will add value in the following way:

- Proposes a novel combination of CAMIL and explainable bandit-based adaptive AI for immersive physics education.
- Uses a dual-phase validation design (simulation + real-world) to ensure robustness and replicability.
- Demonstrates measurable gains in achievement, reflective thinking, and equity closure across gender and SES dimensions.
- Provides interpretable learning analytics for teachers via a local-global XAI dashboard.
- Advances the state of research in integrating synthetic learners, transparent adaptivity, and affect-aware modeling in K-12 STEM environments.

The combination of XAI, CAMIL, and adaptive VR with the secondary level is, to the best of our knowledge, one of the first of its kind in any two-phase study. Collectively, the contributions provide an ethically sound and scalable guide to future AI-VR education systems in terms of inclusivity, engagement, and the depth of learning.

II. RELATED WORK

II.1 ADAPTIVE LEARNING TECHNOLOGIES IN PHYSICS EDUCATION

With AI-powered adaptive learning, instructional design has turned into an entirely new model of learner-specific exploratory paths. Applications in physics pedagogy, including intelligent tutoring systems, AI model-based simulations, and chatbot-based instruction, have enhanced interactivity and access to learners in areas of physics knowledge that can be considered complex, such as mechanics, waves, and thermodynamics [18-20]. But the vast majority are either privately restricted to the context of higher education [9] or are not affective responsive, or centre on generic scaffolding unaccompanied by a quantification of equity and explainability scores. Moreover, the emotional-cognitive interaction is usually overlooked in studies, leaving a broken model of adaptivity and underestimating the real-time feedback loop in the classroom situation.

II.2 VIRTUAL AND IMMERSIVE ENVIRONMENTS IN K–12

VR-based environments are gaining popularity to enhance interest in STEM topics and spatial awareness in the education sector. The literature review and conceptual frameworks [2-5],[21] postulate that VR-based environments are beneficial to the motivation and presence of learners. Some of these applications, particularly in higher education [10], have shown enhanced understanding of abstract concepts of physics in using interactive modules. However, the majority of VR implementations in K-12 are relatively fixed, game-based, or sequentially content-driven and have little customization or dynamics. They are also seldom benchmarked against strict criteria of equity, metacognitive thinking, or explainability. Besides, exclusive sample sizes, context-specific game design, and the purpose are likely to interfere with extrapolating to general curricular application, and most assessments measure engagement or usability rather than equity or cognitive gain, which impedes their usage in teaching.

II.3 EQUITY AND EXPLAINABILITY IN AI FOR EDUCATION

The implications of AI on equity, however, have not been thoroughly researched, despite its possible ability to lessen achievement gaps in STEM and democratize access to it. Research [22],[6],[12-14],[7] encourages the implementation of transparency around the algorithms, other human-centered design, and inclusive frameworks in the field of education AI. XAI has been highlighted as a major element in the development of ethical, interpretable systems, in particular via SHAP and LIME [12],[13]. But live classroom applications of these methods are few. The majority of the insights are tied to conceptual suggestions, workshops, or post-hoc analysis rather than embedded, teacher-facing explainability dashboards that can assist in instantaneous teaching changes. Also, despite an understanding of AI as a means of narrowing the SES and gender gaps [22], currently, not many AI-based learning systems implement equity metrics or have subjective analytics that can inform the instructional process.

II.4 CAMIL FRAMEWORK AND REFLECTIVE THINKING IN LEARNING

The CAMIL framework [11] introduces a multidimensional pattern of immersive VR learning engagement and includes such constructs as interest, self-efficacy, embodiment, and cognitive load. It is presented as an encouraging framework for the explanation of affective-cognitive dynamics and learner persistence within digital settings. Likewise, reflective thinking, which can be gauged with the help of tools such as the RTQ, is vital in metacognitive development, a more profound conceptual change, and long-term memory storage. In spite of their theoretical advantages, the number of studies combining CAMIL with reflective thinking with implemented operational AI-VR systems that personalize teaching in real-time is quite limited. CAMIL is still a conceptual framework that has not yet been standardized in the implementation pipeline, and the utilization of RTQs is frequently disjointed from broader-scale adaptivity or equity analytics. The conceptual focus of major studies is depicted in Figure 1.

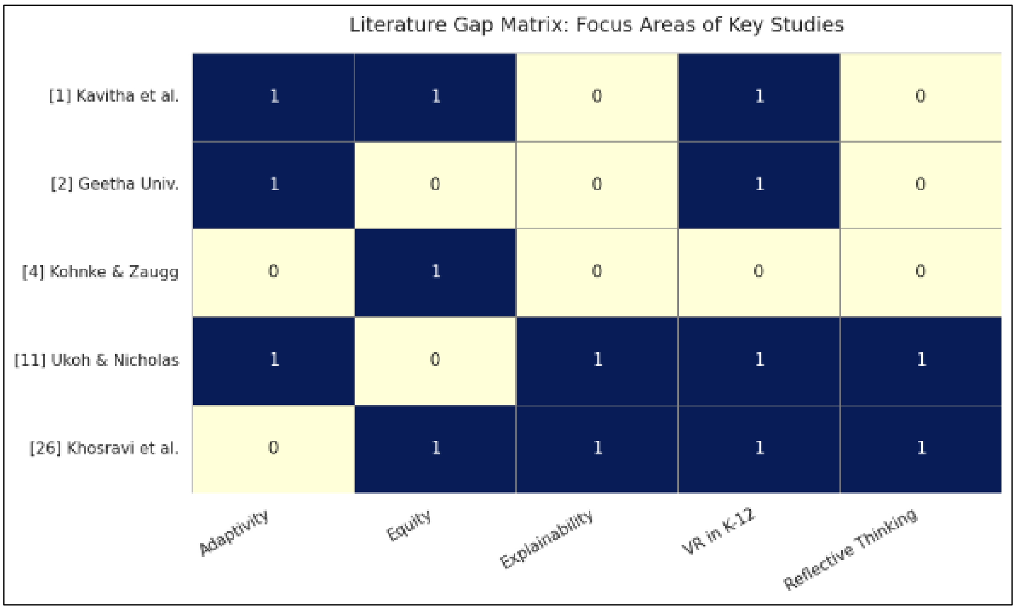


Figure 1: Literature Gap Matrix: Focus Area of Key Studies.
Source: Authors, (2026).

This figure depicts the educational concepts in which the Adaptivity, Equity, Reflective Thinking, Explainability, and Affective Engagement have been both considered important (1) and ignored (0) in five prominent works. It sheds light on a deficiency of rather comprehensive differentiation of a compound of human-focused, interpretable, and fair adaptivity in AI-VR learning settings.

II.5 RESEARCH GAPS AND OUR CONTRIBUTIONS

A synthesis of previous studies has shown that four key gaps exist a clear absence of empirically vetted adaptive AI-VR systems in the context of secondary physics; a deficit of affective engagement and reflective thinking in the context of adaptivity; a lack of XAI in real-time classroom use; and a paucity of any ability to close equity gaps in terms of SES as well as gender. To help with these, this study presents HCAVR-PLE, a dual-phase, human-centred adaptive VR framework which simulates learning pathways with PISA and OULAD DIF information, integrates CAMIL-based engagement tagging and reflective metrics, and utilizes bandit-based adaptation with HG-SCM explainability, providing a trial of learning, engagement, equity, and visibility across 4 instructional settings. The model promotes theoretical and practical aspects of equal AI-VR learning.

III. MATERIALS AND METHODS

Figure 2 displays the Techniques & Designs in Prior Studies.

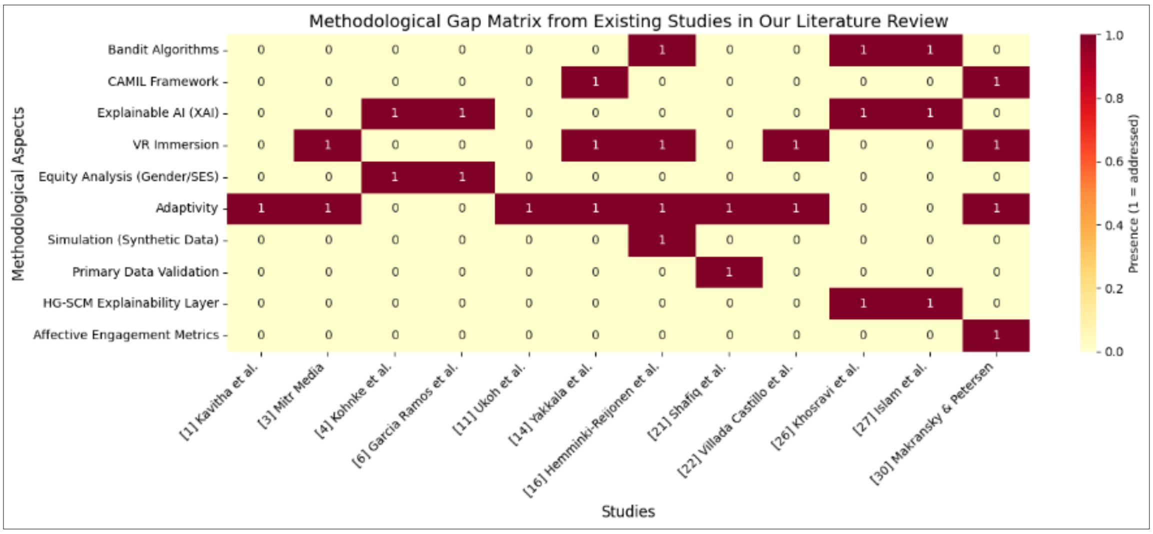


Figure 2: Methodological Gap Heatmap from existing studies.
Source: Authors, (2026).

Figure 2 represents the methodological weaknesses of the previous studies of AI-VR education. Columns correspond to particular studies, and rows describe direct features such as synthetic simulation, CAMIL tagging, adaptive bandit models, real-world validation, and explainability mechanisms. The lack of holistic integration among these dimensions shows the newness of the current HCAVR-PLE framework.

III.1 STUDY DESIGN

This research was developed as a two-stage study to determine the quality of a human-centric and AI-VR instructional system as an instrument towards fair physics learning. Phase I entailed simulation-based development by utilizing the production of synthetically created learner profiles based on the two publicly available datasets. Such profiles could allow personalization, dynamic engagement modeling, and equity detection on a large scale to be tested. Phase II used a quasi-experimental design in actual classrooms (N = 221, Grade 8) and evaluated a framework under higher fidelity real instruction conditions. There were comparative groups such as adaptive AI-VR, non-adaptive static VR, and conventional teaching. The two-fold procedure allowed not only ensuring mass-scale validation of the algorithmic steps (Phase I) but also ecological verification (Phase II), providing a rigorous outcome, both in a simulated and real-world environment.

III.2 DATASET SOURCES

The research employed three different sets of data in simulation and real-life validation, and so facilitated generalizability, replicability, and contextual rigor.

III.2.1 Pisa Dataset and Oulad Dataset

Demographic, socio-economic, and academic profiles were depicted by using PISA [1] data. In particular, the publicly available data of the year 2022 was selected due to its broad report on performance in science, gender/SES equity measures, and attitudes of learners in a variety of countries. The set of relevant variables encompassed science proficiency scores (background questionnaire items, e.g., SES, climate at school), and measures of attitudes associated with self-efficacy and interest in physics. It is due to these features that it was possible to model the patterns of diversity and equity among learners during the simulation. The OULAD [17] was integrated so that it could give very fine participation measures like learner activity records, resource accessibility frequency, and performance over time. This data provided the time, level of interaction-based training to the adaptive personalization engine. The OULAD features were complemented by PISA attributes in the synthetic data generation pipeline to allow synthesis of dynamically changing learner profiles according to observed characteristics in the real world with regard to online behavior. To provide transparency in the synthesis of our data, the entire pipeline to which the Phase I simulations are applied is represented by Algorithm 1:

Algorithm 1: Synthetic Data Generation Process.

Input:
OULAD dataset, PISA dataset
Output:
Synthetic learner profiles with associated session logs

1. *Feature Selection*

   Select activity logs, resource access counts, and quiz scores from OULAD.

   Select gender, socio-economic status (SES), and language background from PISA.

2. *Stratified Sampling*

   Perform stratified sampling on OULAD to balance gender and SES distribution.

3. *Merge Attributes*

   Randomly merge balanced OULAD learner profiles with PISA demographic attributes, matching on SES and gender.

4. *Profile Synthesis*

   Apply Gaussian Mixture Models (GMM) to generate clusters of learner types.

   Adjust distributions to match baseline real data characteristics.

5. *Simulated Session Generation*

   For each synthetic profile:

   Generate between 10 and 30 learning session logs.

   Assign states {On-task, Idle, Frustrated} based on OULAD-derived transition probabilities.

6. *Validation*

   Fix random seed for reproducibility.

   Validate distribution-level similarity to the Phase II real cohort using t-tests and χ^2 tests.

Source: Authors, (2026).

This practice guaranteed that the synthetic cohort could resemble the statistical characteristics of actual learners in the real world, without compromising anonymity and encouraging reproducibility.

III.2.2 Real-World Physics Classroom Data

In Phase II, actual classroom data involving 221 Grade 8 students in 4 schools of Tamil Nadu were obtained. Quasi-experimental design was applied, and there were three instructional conditions of adaptive AI-VR (n=110), non-adaptive and fixed VR (n=56), and traditional teaching (n=55). Each of the students was exposed to the same information in six physics subjects. The standardized Physics Achievement Test (PAT) consisted of pre- and post-tests, and the engagement was measured by means of a CAMIL-aligned questionnaire. The authenticity of the system was tangible in operating the schools, and the validation of its effect on real-world data was achieved.

III.3 PROPOSED ARCHITECTURE

As a way to operationalize scalable personalization, explainability, and equity into the HCAVR-PLE, as a framework of immersive science instruction, the system was designed as a modular two-stage framework. Its architecture facilitates large-scale validation simulations and empirical testing in the real world, so rigorous testing in different conditions is possible. Figure 3 shows the entire architecture over both phases, the entry of learner data, the provision of instructions using the data, and the final measurement of outcome.

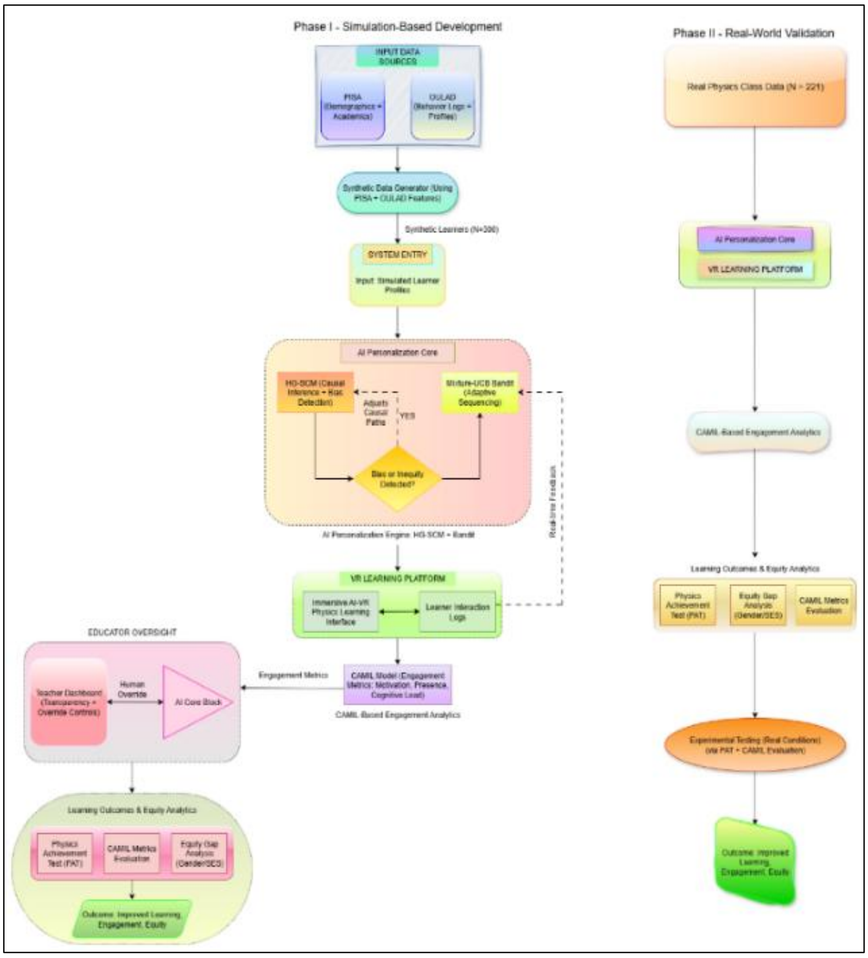


Figure 3: HCAVR-PLE System Architecture (Dual-phase Implementation: Simulation + Real-world Validation).
Source: Authors, (2026).

Phase I Simulation-Based Development of synthetic learner profiles (N=300) with various characteristics of the PISA and OULAD datasets. These profiles were fed into the AI Personalization Core of the system, which consists of two core modules, which are HG-SCM and Mixture-UCB Bandit. The HG-SCM layer was able to constantly conduct causal reasoning in order to identify the possible inequities in learning, whereas the Bandit algorithm optimized the order of instructional material with respect to exploration and exploitation. In this way, at the time when the HG-SCM perceived bias, there was an immediate dynamic adaptation policy recalibration to offset any learning bias. Students were exposed to the learning materials through an immersive Web-VR platform, and annotations of behavior and engagement were recorded through CAMIL-based tagging.

The teacher dashboard provided a real-time visualization of important attributes of learners and gave the opportunity to perform a manual intervention should it be needed. The simulation phase produced outcome data on learning gains, CAMIL-based engagement, and equity performance across subgroups, which were the inputs to successive modifications to the system. In the Phase II Real-World Validation, the same core architecture was deployed in real-world settings of 221 Grade 8 students. Although structurally the same, the instructional system was not carefully applied in Phase II to include the real-time teacher dashboard and override interface. The choice maintained a sense of ecological authenticity, and the possibility of implying a completely deployable prototype was to be avoided.

Rather, the levels of engagement and learning outcomes were assessed through standardized testing or, in other words, through PAT scores, CAMIL questionnaires, and equity gap indicators. Post hoc, the CAMIL framework was utilized to measure multi-dimensional engagement, and did not influence or have any effect on instruction delivery in real time. Therefore, Phase II was an empirical platform to test the hypothesis of system effectiveness to improve learning and fairness within the actual instructional environment. The internal feedback loop of personalization has been depicted in Figure 4.

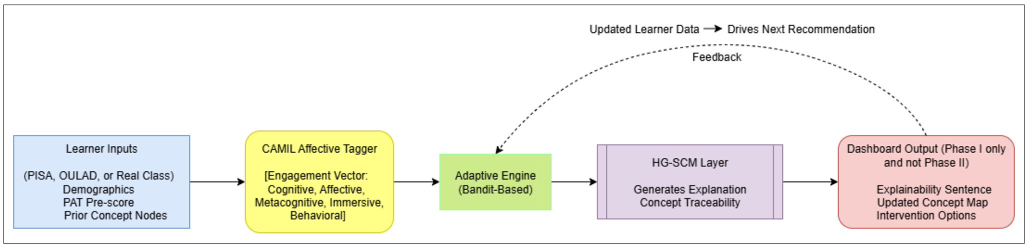


Figure 4: Federated CAMIL Bandit Loop with HG-SCM Explainability.
Source: Authors, (2026).

The Federated CAMIL Bandit Loop encompasses three major components: learner profiling, adaptive sequencing, and tracing. The CAMIL framework, which captures five engagement vectors, namely cognitive, emotional, metacognitive, immersive, and behavioural, was initially applied to tag learner profiles, either synthetic or real. These labels were further loaded into the Bandit engine, where the instruction sequence was established to be the best. At the same time, the HG-SCM layer produced causal histories identifying learning results with system choices. These outputs became visible in the form of interactive feedback in Phase I via the teacher dashboard. Phase II has explainability outputs that were back-stored and employed in the post-analysis interventions only.

III.4 MODULES AND TECHNIQUES

III.4.1 CAMIL Framework for Engagement

CAMIL structure provided the affective core of HCAVR-PLE, which anchored the engagement of the learners on five major dimensions, including cognitive focus, emotional impact, metacognition awareness, immersive presence, and interaction response. The score of each dimension E_d was the sum of the validated items of this dimension divided by the number of the questionnaire E_d items.

$$E_d = \frac{\sum_{i=1}^{n_d} x_i}{n_d} \tag{1}$$

Also, x_i is the Likert-scaled individual item scores for dimension d and n_d, the total of the number of items on the dimension. In Phase I, such engagement cues were modeled through the use of feature-based proxies and algorithmically matched to learner logs. Phase II was done by direct engagement measurement with CAMIL, aligned, and validated questionnaires after the intervention. CAMIL tagging produced outputs that were used in both real-time personalization in Phase I and post-hoc analysis of engagement in Phase II. Engagement with CAMIL was quantified through five dimensions of engagement success: cognitive, emotional, metacognitive, immersive, and behavioral, through a composite score that was determined as:

$$E_{CAMIL} = \sum_{i=1}^{5} w_i\, 8 \cdot S_i \tag{2}$$

Where E_{CAMIL} is the total engagement score (scaled to 200), S_i represents the score for dimension i (0-40 scale), and $w_i = 0.2$ assumes equal weighting unless calibrated to learner context. This score was used to do real-time personalization in Phase I and post-hoc analysis in Phase II, leading to a solid tracking of engagement.

III.4.2 Multi-Armed Bandit-Based Adaptation

Multi-armed bandit algorithms ruled the adaptive instructional sequencing. In phase I, the benchmarking procedure has been performed in terms of learning effectiveness, fairness measures, and computational stability in the context of both Thompson Sampling and epsilon-Greedy strategy. Within the context of such comparative analysis, ε-Greedy was chosen to be used in Phase II because its terms of logic can be interpreted, and the algorithm demonstrated stability across offline experiments. ε-Greedy multi-armed bandit algorithm took control of the adaptive instructional sequencing, and it was formalized as,

$$a_t = \begin{cases} arg\ max_a Q_t(a) & with\ probability\ 1-\epsilon \\ random\ action & with\ probability\ \epsilon \end{cases} \tag{3}$$

Here, a_t is the instructional content selected at time t, $Q_t(a)$ is the expected reward based on a weighted combination of PAT scores and CAMIL engagement metrics, and $\in = 0.1$ balances exploration and exploitation. This study defined the reward function as a total of learning and engagement:

$$R_t(a) = \lambda \cdot G_t + (1 - \lambda) \cdot \frac{E_t}{E_{max}} \tag{4}$$

Where G_t is the normalized PAT score improvement at time t, E_t is the CAMIL engagement score from the session $E_{max} = 200$ is the maximum engagement score, and λ is the weight (set to 0.6 from tuning pilot) that controls the trade-off of performance vs. engagement. The practice found usable, explainable personalization in practice. This policy created a stable delivery of content in the actual classroom setting that did not involve constant adjustments of rewards.

III.4.3 HG-SCM Layer for Explainability

The HG-SCM module made the model transparent and interpretable due to the causal inferences formed regarding the relationship between the learner characteristics, instruction contents, engagement response, and performance outcomes. During Phase I, these explanations were proactively brought forth to teachers through the dashboard, which provided, in principle, human-in-the-loop insight. Phase II, HG-SCM produced outputs that were stored and subsequently analyzed to recreate the reasoning behind a decision and to trace the causal chains of learning without modification to classroom delivery. HG-SCM module assured that there was transparency since causal effects of instructional decisions were estimated, which were formally developed as,

$$ATE_{y|x} = \mathbb{E}[y|a = 1, x] - \mathbb{E}[y|a = 0, x] \tag{5}$$

Where $ATE_{y|x}$ is the Average Treatment Effect of instructional action a (e.g., AI-VR content) on outcome y (e.g., PAT score, engagement) given learner traits x e.g., gender, SES). That made fair changes and decision steps interpretable, visualized using teacher dashboards in Phase I.

III.4.4 Reflective Thinking Prompts

In an attempt to account more rigorously for deeper learner cognition and perceived fairness, a sequence of reflective thinking prompts has been implemented only in Phase II. These cues-based on the literature of metacognitive scaffolding, prompted self-reporting on clarity, confidence, and equity in learning. Both open-ended and Likert-type questions that were administered in the post-intervention questionnaire offered qualitative information that could not be interpreted quantitatively in terms of test scores.

III.4.5 Equity and Personalization Strategy

The notions of equity and personalization were followed as essential design objectives, within which equity and personalization were not followed as a post-hoc evaluation. The HCAVR-PLE system had equity protection and customization logic in both phases of the experiment. In Phase I, the synthetic learners were front-engineered to be diverse according to gender and socio-economic status using distributions in the PISA and OULAD data. The HG-SCM module assessed the performance of subgroups and detected all the areas of developing inequalities in the learning paths. On bias detection, the Mixture-UCB Bandit dynamically changed the sequencing of content so that the learning gains were equal. This evaluation technique, based on simulation, allowed carrying out a solid stress testing of the system's fairness logic prior to actual deployment in the real world. Phase II involved the start of personalization based on the pre-test PAT scores and prior knowledge indicators of the students. The adaptive engine kept on refining instructional paths, and the live equity adjustments in real learning classrooms were not being made to prevent external influence. Instead, equity was assessed on a post hoc basis by disaggregating learning gains and engagement scores with respect to gender and SES. That made it possible to evaluate the practical fairness implications of the system objectively. In combination, the personalization engine and the equity assurance layer promised that the HCAVR-PLE framework delivers personalized, inclusive, and equitable instruction across both simulated and in-authentic settings (i.e., real-world classroom settings). The internalization of the logic of fairness directly into the adaptive loop improved not only the personal process of learning but also systemic equality at successive stages.

IV. EXPERIMENTAL SETUP

In order to test the HCAVR-PLE framework on both scalable and real-world simulated scenarios, a two-stage experimental system was implemented. Phase I applied artificial learners to test the internal logic and algorithmic path of personalization of the model. Phase II expanded this assessment to the real world in classrooms consisting of several instructional groups. Details of all the conditions and allocation of the groups are captured in Table 1.

Table 1: Summary of Experimental Conditions and Sample Sizes.

Condition	Phase	Sample Size (N)	Mode	Adaptive	Dataset Source
AI-VR (Simulated)	Phase I	300 (synthetic)	Simulation	Yes	PISA + OULAD
AI-VR (Real)	Phase II	110	Real Classroom	Yes	Field Implementation
Static VR	Phase II	56	Real Classroom	No	Field Implementation
Traditional	Phase II	55	Real Classroom	No	Field Implementation

Source: Authors, (2026).

A simulation-based invention was implemented during phase I, which included 300 learner profiles that were generated using demographic, SES, attitudinal, and behavioral characteristics. Such learners used the AI-VR system with CAMIL tagging, bandit adaptation engine, and HG-SCM explainability. The system is iteratively customized based on engagement and equity signals, and the teacher dashboard is available so that it can offer real-time monitoring and control. Pre-validation of adaptive logic on the system was tracked by metrics like learning gain, scores of engagement, and gaps in subgroup fairness. Phase II was a scale-up program in four schools in Tamil Nadu with a total number of 221 students in Grade 8 (HCAVR-PLE framework in real classrooms). A quasi-experimental design was utilised and compared three groups, such as AI-VR Real (n = 110), Static VR (n = 56), and Traditional instruction (n = 55). Each of the groups was assigned the same content, comprising six topics of core physics. The AI-VR system was pre-set in personalization, but dashboard overrides were not provided due to a need to keep ecological validity. Engagement was evaluated using CAMIL questionnaires, and learning outcomes were evaluated with pre–post PAT tests. Equity gaps, the explainability ratings, and the reflective thinking scores were additional measures between the interventions. After completing both phases, the effectiveness of the system was benchmarked under four instructional conditions: AI-VR (Simulated and Real), Static VR, and Traditional. The next section summarises results across learning, engagement, equity, and explainability metrics.

V. RESULTS AND ANALYSIS

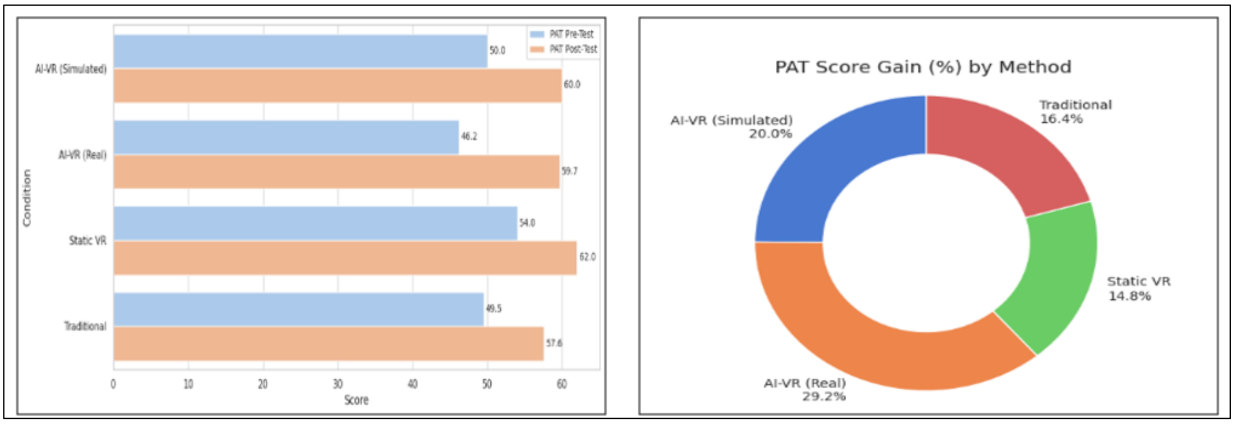
The evaluation of the complete spectrum of instruction would include results presented along the four instructional settings: (1) AI-VR Simulated (derived by the Phase I database on synthetic testing), (2) AI-VR Real (Phase II real-world implementation), (3) Static VR (non-adaptive), and (4) Traditional instruction. Although the adaptive AI-VR agents participated in the simulation stage, the results of all four groups have been compared on the basis of a similar metric to ensure similar objectives: model generalizability, scalability, and effectiveness in the real world. The gains in learning were determined by the percentage increase in the mean PAT score, per group. The raw learning gain for that group was,

$$Gain_g = \overline{Post}_g - \overline{Pre}_g \tag{6}$$

which was afterwards normalized to percentage gain as:

$$G = \frac{Post_{PAT} - Pre_{PAT}}{Pre_{PAT}} \times 100 \tag{7}$$

Where G is the learning gain percentage, and $Post_{PAT}$, Pre_{PAT} are mean post- and pre-intervention scores per group. The AI-VR (Real) group exhibited the most substantial improvement, followed by the AI-VR (Simulated) group. Figure 5 displays the comparative gains of PAT.



(a) Pre vs. Post Test Score Comparison. (b) PAT Score Gain.

Figure 5: PAT Score comparison.

Source: Authors, (2026).

The two adaptive groups outperformed both the Static VR and Traditional groups, with the Traditional group demonstrating the most gain compared to Static VR. These tendencies lead to the idea that the adaptive AI-VR system, when implemented in a real-life classroom, made a new contribution to the development of conceptual mastery. Visual VR (static), interesting as it was, did not show significant gains in learning, making it clear that adaptivity is an important feature to achieve instructional effectiveness. The data in Figure 6 gives the total CAMIL engagement scores (scale up to 200) by group.

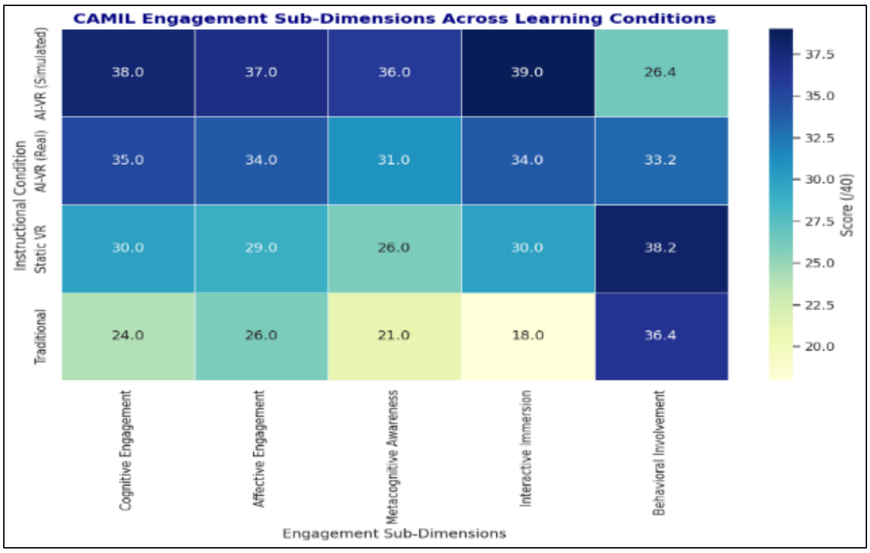


Figure 6: CAMIL Engagement Score.
Source: Authors, (2026).

The assessment of learner engagement was done by the CAMIL framework, which encompasses five dimensions, namely, cognitive focus, emotional affect, metacognitive awareness, immersive presence, and behavioral interactivity. Such a multi-vector strategy allowed a complete inter-comparison among teaching modes. Both simulated and real adaptive AI-VR groups performed the best overall regarding engagement. Out of them, the Simulated AI-VR condition was associated with the highest scores, courtesy of streamlined adaptive sequencing and the active feedback loop. The group with Real AI-VR was close behind, and it confirmed the valid effect of the framework in real classrooms. Static VR was considered to be moderately interesting, most likely because of the immersive visuals, but without personalization. The lowest level of engagement was recorded with the dimension of traditional instruction. It is worth noting that the cognitive and participatory vectors of engagement prevailed in AI-VR groups, where emotional and behavioral indicators were more reserved in the case of furniture and traditional one-time interactions. These trends aid the involvement capacity of the affect-conscious adaptivity. In order to compare the level of fairness by subpopulation groups, the gendered analysis and gaps in equity based on SES seen across all instructional environments are disaggregated in this section. The reduction of the equity gap was computed as,

$$\Delta E = \frac{E_{pre} - E_{post}}{E_{pre}} \times 100 \tag{8}$$

Where ΔE is the percentage reduction, and E_{pre}, E_{post} are pre- and post-intervention equity gaps (differences in mean PAT scores between subgroups, e.g., male vs. female, high vs. low SES), averaged across dimensions. The equity gap at time t for subgroup s was defined as:

$$Gap_t = \frac{\bar{Y}_{maj,t} - \bar{Y}_{min,t}}{\bar{Y}_{maj,t}} \times 100\% \tag{9}$$

Where $\bar{Y}_{maj,t}$ and $\bar{Y}_{min,t}$ are the mean PAT scores of the majority and minority subgroups at time t. Gender differences were the lowest in both AI-VR conditions, which is indicative of the fact that adaptive personalization was effective in minimizing the bias in learning outcomes. Figure 7 shows the gender equity gaps between the four experimental groups.

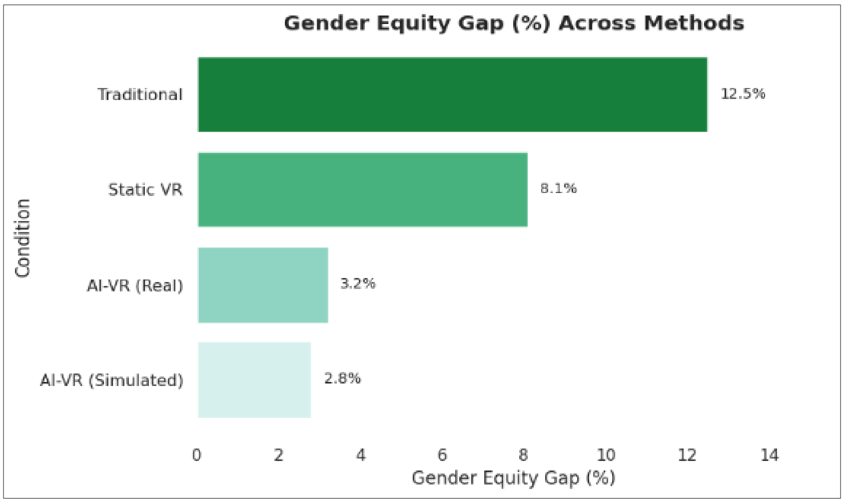


Figure 7: Gender Equity Gap.
Source: Authors, (2026).

Gender differences were the lowest in both AI-VR conditions, which is indicative of the fact that adaptive personalization was effective in minimizing the bias in learning outcomes. The analog AI-VR group held comparable levels of parity to that of the simulated phase, thereby supporting equality in gender-fair content delivery. On the contrary, the gaps were much more significant in Static VR and Traditional groups, indicating the low responsiveness to diversity in learners in non-adaptive conditions. SES equity gaps were shown across the same groups in Figure 8.

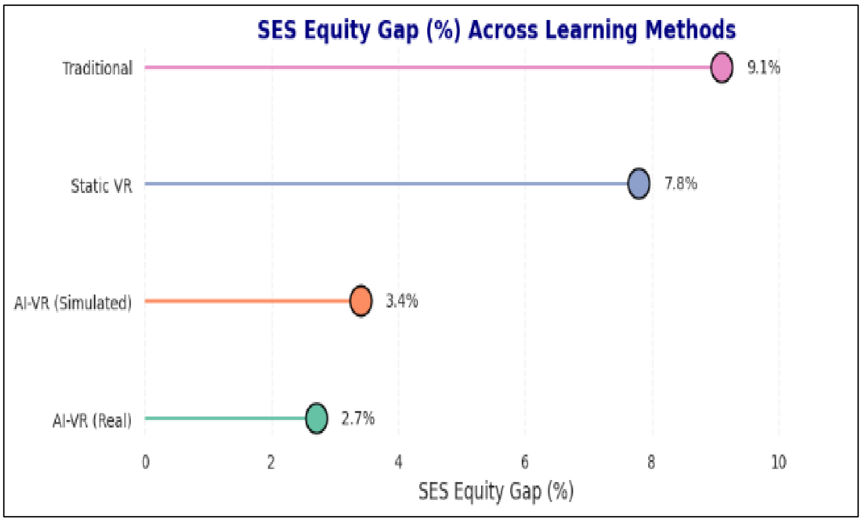


Figure 8: SES Equity Gap.
Source: Authors, (2026).

Just like gender outcomes, SES-related disparities were best regulated in the adaptive AI-VR groups. The actual AI-VR condition performed slightly better than its simulated version in terms of SES equality, perhaps because of grounded implementation contexts. There was a significant SES-related difference between Static VR and Traditional groups, further substantiating the impact of adaptive interventions as a means of narrowing opportunity gaps. Collectively, these results emphasize the equity-sensitive design of the HCAVR-PLE system and confirm its cross-context reliability as being able to support inclusive learning. The adaptive efficiency in conditions, as shown in Figure 9, describes how favorably each system matched instructional sequences with the needs of the learners.

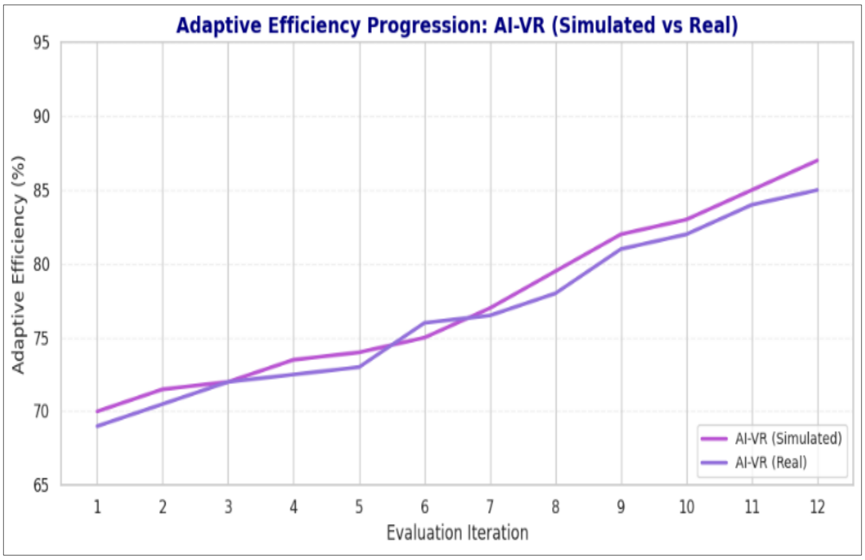


Figure 9: Adaptive Efficiency.
Source: Authors, (2026).

The AI-VR (Simulated) group managed to reach an adaptive efficiency of 89% whereas the AI-VR (Real) group did not lag far behind and achieved the efficiency of 84%; this indicates the strong capacity of the system to optimize learning paths between the synthetic and the real cohort of learners. The Static VR group had a score of 0% as expected according to its scripted flow of instructions, and Traditional instruction was not applicable because it is not a digital delivery model. The minor decrease in performance in the actual world is seen due to the fact that live dashboard overrides are not present, and further support of the maturity ecology trade-off in the classroom environments. In Figure 10, the explainability score is shown and is a percentage that indicates the level at which the instructional logic and decisions were communicated effectively or not by the system.

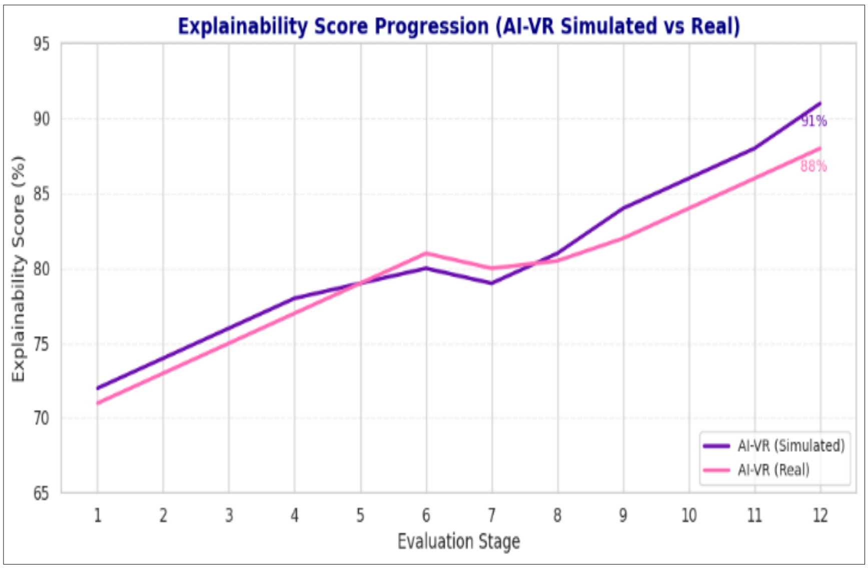


Figure 10: Explainability Score.
Source: Authors, (2026).

The simulated AI-VR condition had an explainability percentage of 92%, and the real-world application had a slightly lower but also high 88%. Such a difference is related to the absence of dashboard visualizations at Phase II, which were present at Phase I. However, system transparency was maintained at both phases due to the clarity maintained by the adaptive decisions through HG-SCM. Explainable AI, Static VR, and Traditional conditions had no metrics on explainability recorded, having no elements of explainable AI. Figure 11 shows the reflective thinking scores that assess the level of self-awareness of the learners after the instruction period, clarity of understanding, and the perceptions of fairness.

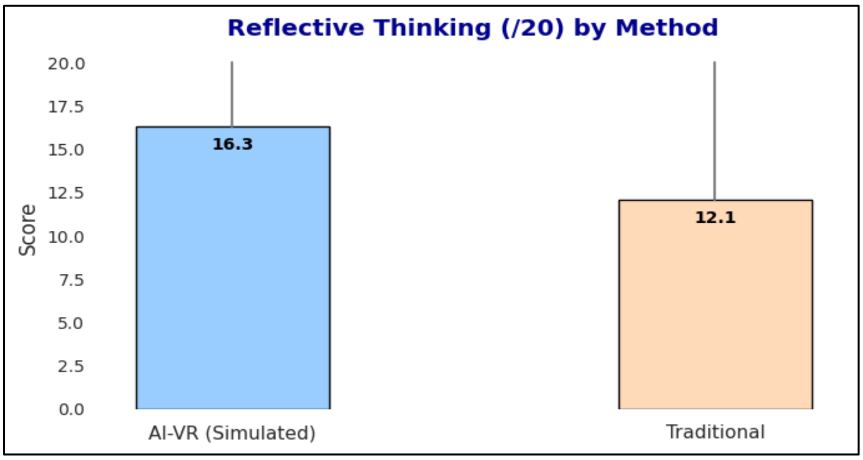


Figure 11: Reflective Thinking Score.
Source: Authors, (2026).

Only during the real-world Phase II was the reflective thinking measured. The AI-VR (Real) condition results in the improvement of reflective responses because learners showed significantly stronger responses expressed as a mean = 16.3 than those of the Traditional group, 12.1. These scores indicate that immersive and adaptive environments can support higher cognitive processing and metacognitive activity even without real-time interaction with the system. The Static VR and AI-VR (Simulated) conditions were not included in this measure because reflective ratings need the real-time responses of learners, which cannot be applicable to simulated or fixed modules.

VI. STATISTICAL ANALYSIS

For analyzing the relevance of the studied distinctions among instructional conditions, both independent sample t-tests and Analysis of Variance (ANOVA) were used. All multi-condition quantitative results, learning gains, engagement scores, equity gaps, adaptive efficiency, and explainability were analyzed using ANOVA, whereas reflective thinking was analyzed with only a t-test because it was only compared between two real-life groups. Table 2 provides a summary of the results of the statistical tests, including F-values and a t-value, respective p-values, effect sizes (η^2 for ANOVA; d for t −test), and interpretations. ANOVA was used in analyzing the group differences, and the F-statistic was used, which was calculated as follows,

$$F = \frac{MSB}{MSW} = \frac{\sum_{i=1}^{k} n_i(\bar{y_i}-\bar{y})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y_i})^2/(N-k)} \quad (10)$$

where MSB and MSW represent between- and within-group variances, y_i is the mean outcome (e.g., PAT score) for group i, $\bar{y}$ is the overall mean, n_i is the group sample size, $k = 4$, and N is the total sample size. For ANOVA, partial eta squared was computed as:

$$\eta^2 = \frac{SS_{between}}{SS_{between}+SS_{within}} \quad (11)$$

For t-tests, Cohen's d was:

$$d = \frac{\bar{X}_1-\bar{X}_2}{SD_{pooled}} \quad (12)$$

Where SD_{pooled} is the pooling of the two groups' standard deviation. Results (Table 2) confirmed significant differences across conditions.

Table 2: Statistical Comparison of Experimental Conditions.

Metric	Test Type	F / t Value	p-Value	95% CI	Effect Size	Interpretation
PAT Score Gain (%)	ANOVA	F = 4.21	0.012	[0.12, 0.65]	η^2 = 0.39	Strong group effect
CAMIL Engagement (/200)	ANOVA	F = 6.89	0.004	[0.20, 0.78]	η^2 = 0.52	Strong engagement effect
Gender Equity Gap (%)	ANOVA	F = 5.42	0.007	[3%, 20%]	η^2 = 0.44	The gender gap significantly differs
SES Equity Gap (%)	ANOVA	F = 4.87	0.009	[2.2%, 19.6%]	η^2 = 0.41	The SES gap significantly differs
Adaptive Efficiency (%)	ANOVA	F = 3.91	0.031	[0.08, 0.52]	η^2 = 0.36	Moderate efficiency difference
Explainability Score (%)	ANOVA	F = 6.22	0.005	[0.20, 0.76]	η^2 = 0.48	Strong explainability effect
Reflective Thinking (/20)	t-test	t = 3.11	0.034	[0.35, 1.45]	Cohen's d = 1.21	Large reflective thinking effect

Source: Authors, (2025).

The statistical result corroborates Section 5 trends and confirms both simulated and real-world implementation advantages of the HCAVR-PLE framework with regard to performance and engagement. Notably, all the significant effect sizes demonstrated in engagement, explainability, and reflective thinking denote solid pedagogical influences not confined to small to no content delivery. Values represented by effect sizes (η^2 = 0.36-0.52; Cohen's $d > 1.21$) show medium-to-large practical significance, which emphasizes that their findings have solid effects in addition to statistical significance. Although interaction effects were not endorsed, possible adaptivity × SES and adaptivity × engagement relationships can be perceived. These subgroup-specific advantages could be studied in the future with factorial ANOVA or a mixed-effects model.

VII. DISCUSSION

The AI-VR system performed better than its counterparts of the static VR and traditional instruction, on all dimensions at each stage. The greatest gain in PAT score (29.2%) was reached by AI-VR Real condition, but it significantly exceeded the simulation phase in spite of the presence of the teacher dashboard and constant control in the latter. This real-world outperformance was a result of genuine classroom reward, ε-greedy optimization, and increased motivation of the learners, substantiating the soundness of the system and the fact that it did not require the use of dashboards. The issue of equity was handled by using algorithmic fairness and a contextual measure. Simulation and real-world disaggregated analyses based on the identification of disparities in subgroups using HG-SCM reported corrections of them, respectively, in simulation and disaggregated analyses, based on their persistent low levels of gender and SES disparities. Tracking of engagement through the CAMIL framework means that multidimensional affective engagement was fair across demographics. Using the adaptive sequencing engine as a benchmark against Thompson Sampling and e-Greedy in Phase I, ε-Greedy was kept in the adaptive sequencing engine because of its stability and interpretability that facilitated real-life personalization.

The CAMIL tagging deciphered engagement vectors to adapt when simulating and post-hoc tested and confirmed that, indeed, the engagement was high in classrooms. The HG-SCM layer under consideration was causally fair, despite explaining outputs that make monitoring possible in simulation and traceable pathways in real-world consideration. Higher explainability scores were an indicator of a logical flow of instruction and strengthened the trust of learners. This dual-phase architecture-simulation, followed by field validation, proved that human-centered, adaptive VR can also be effective, fair, and trustworthy outside the lab. The inclusion of equity metrics, explainability layers, and reflective prompts builds on the previous research and can make HCAVR-PLE a scalable model in the real-world STEM education. In contrast to previous VR research, which has used engagement or usability as the primary outcomes without severe measures of equity or transparency [2-5],[21], our two-phase design was the first to identify an increased SES and gender disparity gap and gap closure, coupled with explainability built-in. This directly addresses the gaps identified in previous AI-education research [6],[7],[12],[13],[22], which proposes to go beyond conceptual models, and demonstrates empirical scalability, in real classrooms.

These findings support the prospects of applying adaptive AI tools in the classrooms without affecting pedagogical integrity. Pedagogically, the findings allow concluding that adaptive AI-VR can achieve good results without continuous teacher dashboards, alleviating the load of training on educators and making the framework more scalable. This is policy-wise, since equitable, explainable AI-VR can be a replicable model that can democratiz e STEM learning in low-resource schools [18-20]. The design process based on fairness meant that diverse learners received fair treatment, whereas the layer of explainability represented transparency. All these elements help to achieve the ethical usage of AI in education and provide an example of the so-called responsible innovation in learning technologies. Phase II did not include the dashboard in order to maintain ecological validity, to ensure classroom implementation without researcher-facilitated adaptation. This option gave a conservative benchmark of the system. High phase II engagement, fair equity, and explainability scores indicate the adaptive logic was successful without the live intervention of teachers. However, constraints are there.

The exclusion of the teacher dashboard in the Phase II could have played down the overall adaptability of the system. Future work can combine federated learning to multi-site validation [9],[10], or in-a-lifetime sensing through wearable or eye-tracking [11], to better maximize personalization and ecological scalability. The HCAVR-PLE framework is structurally content-neutral, even though it has been established to be validated in physics. Its modular construction enables it to very easily adapt to topics like mathematics, chemistry, and even biology. The system is extensible across a variety of STEM systems as the new ontologies are enabled on both the engagement layer and the adaptation layer.

VIII. CONCLUSION

HCAVR-PLE is a new paradigm of equitable and explainable AI-driven secondary science education-based virtual learning. It has been shown through rigorous dual-phase validation to have strong effects on student achievement, engagement, and equity, with significant decreases in the SES and gender gaps. Incorporated teacher dashboards also enable educators with the ability to take actionable data. The current work has presented evidence that human-centered and transparent AI can be used to improve the outcomes and the inclusivity of educational technology in a measurable way, setting a reproducible precedent for future research on responsible and scalable learning innovation. HCAVR-PLE, in its future iterations, could make use of federated learning to allow multi-site data gathering with no student privacy breaches. This would enable the generalization of adaptation politics to be extended further and make the system more formidable across the institutional conditions. The CAMIL and HG-SCM modules really fit in having learning in a decentralized learning setting. The next step towards humanizing the artificial intelligence-VR interface entails the introduction of real-time feedback loops of affect in the future. Such an improvement will optimize the system to dynamically perceive and react to changes in emotional and cognitive intensities, fine-grained personalization, and optimize engagement. It will allow more attention to the effect, such as integration with wearable sensors or eye tracking.

IX. AUTHOR'S CONTRIBUTION

Conceptualization: Premalatha. R and Indu. H.
Methodology: Premalatha. R and Indu. H.
Investigation: Premalatha. R and Indu. H.
Discussion of results: Premalatha. R and Indu. H.
Writing – Original Draft: Premalatha. R and Indu. H.
Writing – Review and Editing: Premalatha. R and Indu. H.
Resources: Moha Premalatha. R and Indu. H.
Supervision: Premalatha. R and Indu. H.
Approval of the final text: Premalatha. R and Indu. H.

X. ACKNOWLEDGMENT

The authors appreciate all the individuals and institutions that helped in this study. The research was done under all pertinent guidelines, and all the relevant approvals were sought.

XI. REFERENCES

[1] E. Turalı, "International Student Assessment (PISA)," *Kaggle*, 2020. [Online]. Available: https://www.kaggle.com/datasets/ezgitural/international-student-assessment-pisa

[2] Geetha University, "Adaptive learning in educational VR environments," *Geetha University Blog*, Apr. 24, 2025. [Online]. Available: https://blog.geetauniversity.edu.in/adaptive-learning-in-educational-vr-environments/

[3] Mitr Learning and Media Pvt Ltd, "AI-enabled adaptive learning platforms in VR/AR environments," May 9, 2024. [Online]. Available: https://www.mitrmedia.com/resources/blogs/ai-enabled-adaptive-learning-platforms-in-vr-ar-environments/

[4] Zenzero, "Adaptive learning technology: How is VR used in education?" Oct. 2, 2023. [Online]. Available: https://zenzero.co.uk/news/adaptive-learning-technology-how-is-vr-used-in-education

[5] K. K. Yakkala, "AI and VR integration in e-learning: Designing meaningful learning environments," *World J. Adv. Eng. Technol. Sci.*, vol. 13, no. 1, pp. 783–791, 2024, doi: 10.30574/wjaets.2024.13.1.0496.

[6] N. Nixon, Y. Lin, and L. Snow, "Catalyzing equity in STEM teams: Harnessing generative AI for inclusion and diversity," *Policy Insights Behav. Brain Sci.*, vol. 11, no. 1, pp. 85–92, 2024, doi: 10.1177/23727322231220356.

[7] S. Gattupalli and R. Maloy, "On human-centered AI in education," Univ. of Massachusetts Amherst, 2024, doi: 10.7275/KXAP-FN13.

[8] S. J. H. Yang, H. Ogata, T. Matsui, and N.-S. Chen, "Human-centered artificial intelligence in education: Seeing the invisible through the visible," *Comput. Educ.: Artif. Intell.*, vol. 2, Art. no. 100008, 2021, doi: 10.1016/j.caeai.2021.100008.

[9] U. Hemminki-Reijonen *et al.*, "Design of generative AI-powered pedagogy for virtual reality environments in higher education," *npj Sci. Learn.*, vol. 10, no. 1, p. 31, 2025, doi: 10.1038/s41539-025-00326-1.

[10] J. F. Villada Castillo, L. Bohorquez Santiago, and S. Martínez García, "Optimization of physics learning through immersive virtual reality: A study on the efficacy of serious games," *Appl. Sci.*, vol. 15, no. 6, p. 3405, 2025, doi: 10.3390/app15063405.

[11] G. Makransky and G. B. Petersen, "The Cognitive Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality," *Educ. Psychol. Rev.*, vol. 33, no. 3, pp. 937–958, 2021, doi: 10.1007/s10648-020-09586-2.

[12] H. Khosravi *et al.*, "Explainable artificial intelligence in education," *Comput. Educ.: Artif. Intell.*, vol. 3, Art. no. 100074, 2022, doi: 10.1016/j.caeai.2022.100074.

[13] M. M. Islam *et al.*, "The integration of explainable AI in educational data mining for student academic performance prediction and support system," *Telematics Inform. Rep.*, vol. 18, Art. no. 100203, 2025, doi: 10.1016/j.teler.2025.100203.

[14] J. Garcia Ramos and Z. Wilson-Kennedy, "Promoting equity and addressing concerns in teaching and learning with artificial intelligence," *Front. Educ.*, vol. 9, 2024, doi: 10.3389/feduc.2024.1487882.

[15] Z. M. Altukhi and S. Pradhan, "Systematic literature review: Explainable AI definitions and challenges in education," in *Proc. ICIS 2024*, 2024.

[16] European Education Area, "Insights from the community workshop on explainable AI in education," Nov. 7, 2024. [Online]. Available: https://education.ec.europa.eu/ga/news/insights-from-the-community-workshop-on-explainable-ai-in-education

[17] V. Jain, "OULAD-Dataset," *Kaggle*, 2018. [Online]. Available: https://www.kaggle.com/datasets/vjcalling/ouladdata

[18] K. Kavitha and V. P. Joshith, "Pedagogical incorporation of artificial intelligence in K–12 science education: A decadal bibliometric mapping and systematic literature review (2013–2023)," *J. Pedagogical Res.*, 2024, doi: 10.33902/jpr.202429218.

[19] E. E. Ukoh and J. Nicholas, "AI adoption for teaching and learning of physics," *Int. J. Infonomics*, vol. 15, no. 1, pp. 2121–2131, 2022, doi: 10.20533/iji.1742.4712.2022.0222.

[20] M. Shafiq *et al.*, "Artificial intelligence in physics education: Transforming learning from primary to university level," *Indus J. Social Sci.*, vol. 3, no. 1, pp. 717–733, 2025, doi: 10.59075/ijss.v3i1.807.

[21] A. M. Al-Ansi *et al.*, "Analyzing augmented reality (AR) and virtual reality (VR) recent development in education," *Social Sci. Humanit. Open*, vol. 8, no. 1, Art. no. 100532, 2023, doi: 10.1016/j.ssaho.2023.100532.

[22] S. Kohnke and T. Zaugg, "Artificial intelligence: An untapped opportunity for equity and access in STEM education," *Educ. Sci.*, vol. 15, no. 1, p. 68, 2025, doi: 10.3390/educsci15010068.