Journal of Applied Engineering and Technological Science
Vol 7(1) 2025:414-438

THE IMPACT OF VIRTUAL LABORATORIES ON STUDENT MOTIVATION AND ACADEMIC PERFORMANCE: AN INTEGRATED FUZZY-SEM AND MACHINE LEARNING STUDY

Tabriz Osmanli
Department of Artificial Intelligence Technology, National Aviation Academy, Baku, Azerbaijan
tosmanli@naa.edu.az

Received: 01 September 2025, Revised: 22 November 2025, Accepted: 27 November 2025
**Corresponding Author*

ABSTRACT
This study explored the impact of virtual laboratories (VLs) on university learning and seeks to fill a gap in the literature: most VL research reports positive outcomes, but rarely explains why they occur or whether psychological mechanisms generalize predictively. The solution comes from a synthetic model combining Fuzzy-SEM, which is great for modelling uncertainty within Likert-based motivation and engagement constructs, with supervised machine learning models that provide causal explanation combined with predictive validation. We analyzed data from 432 undergraduates combining VL usage logs, motivation–engagement surveys, and official academic records. Fuzzy-SEM confirmed a mediated motivation–engagement–performance pathway, which confirms that VLs significantly boost performance primarily by converting motivational activation to sustained engagement. Predictively, the 1D CNN better fitted the classical ML models (AUC-ROC = 0.94) suggesting the possibility of early identification of at-risk students through behavioural and affective proxies. Practical implications should be to apply VLs as complementary motivational approaches to training practice and to monitor prediction weekly for intervention. In theory, the study bolsters engagement frameworks by elucidating how VLs exert their effect. Methodologically, it presents an integrated Fuzzy-SEM + ML pipeline that facilitates both explanatory context and potentially deployable prediction, although it recognizes the limitation of single-institution and self-report.
Keywords: *Virtual Laboratories, Fuzzy-SEM, Machine Learning, Motivation, Academic Performance, Engagement, Predictive Analytics, E-learning*

1. Introduction

Virtual and immersive technologies have revolutionized higher education by enabling easily scalable, safe, and flexible alternatives to the traditional laboratory setting. Although numerous studies have shown that virtual laboratories (VLs) positively influence motivation, engagement, and academic performance, the mechanisms through which VLs exert these effects remain insufficiently explained. Although the existing studies mostly rely on comparison of outcomes, they do not reveal what goes on in motivation and behavioral processes in technology-mediated contexts. Therefore, though the positive trends are consistently observed, the role of VLs remains partially comprehended at the level of cause and psychology.

Recent evidence (2020–2025) describes a number of important patterns. Meta-analyses and quasi-experimental studies have confirmed that VLs dramatically improve student motivation, engagement and conceptual understanding in STEM subjects and are best used in tandem with rather than in lieu of physical laboratories (Li & Liang, 2024; Dodevska et al., 2025; Mahendra, 2025). Concurrently, existing motivation and engagement theories—ARCS, Self-Determination Theory, and multidimensional engagement models—demonstrate the relationship between affective and behavior facets of learning in the context of digital learning environments and the significance of motivational pathways in technology-enabled contexts (Lampropoulos & Kinshuk, 2024).

Simultaneous research in the area of learning analytics suggests that machine learning (ML) algorithms trained on behavioral interaction data can give the ability to predict at-risk students and represent sophisticated learning behaviors with remarkable accuracy (Arévalo-Cordovilla & Peña, 2024; Al-Alawi et al., 2023; Zhao et al., 2025). Methodologically, recent empirical work in psychometrics and SEM indicates that Fuzzy-SEM is ideal for modelling Likert-type constructs in which fuzzy membership functions are better able to understand uncertainty and language variability as compared to the standard SEM (Ringle et al., 2023).

However, many research gaps still need to be filled. First, previous researches rarely explore the motivational mediation pathway (motivation → engagement → performance), leaving it ambiguous whether VLs affect performance via psychological pathways. Second, while VLs create extensive behavioral logs, few studies combine objective usage data with motivational-related constructs in a holistic analytic approach. Third, little research combines causal modeling (through SEM or Fuzzy-SEM) and predictive modeling (through ML), such that explanatory validity and predictive generalizability are rarely compared.

To address these gaps, the present study proposes an integrated approach integrating Fuzzy Structural Equation Modeling (Fuzzy-SEM) and machine learning. Fuzzy-SEM is employed because constructs like the Likert's are based on linguistic uncertainty and fuzzy membership functions yield better causal estimates than regular SEM. However, ML models are implemented since SEM alone cannot be used to represent nonlinear behavioral patterns, nor to predict with high precision the behavior characteristics required to detect struggling learners at an early stage. This approach explains causal pathways as well as predictive performance and therefore provides a comprehensive framework for understanding effectiveness of VL.

Accordingly, this study pursues four research objectives:

1. To model the causal relationships among motivation, engagement, and academic performance in VL contexts using Fuzzy-SEM.
2. To examine whether engagement mediates the effect of motivation on academic performance.
3. To integrate VL behavioral usage logs with survey-based affective constructs.
4. To evaluate the predictive performance of LR, RF, SVM, k-NN, and CNN models for identifying at-risk students.

2. Literature Review

To provide a clear theoretical and analytical explanation, this review first provides a theoretically grounded explanation of how motivation, engagement, and cognitive processing work in the context of technology-enhanced and virtual laboratory (VL) environments. Unlike most of the previous descriptive treatments in the literature, this section explicitly connects theory to the mechanisms tested in the study, so that the narrative could be seen as analytical rather than purely narrative. These theories include Self-Determination Theory (SDT), the ARCS motivational framework, the Technology Acceptance Model (TAM), and the Cognitive Theory of Multimedia Learning (CTML). Together, they show why virtual laboratories (VLs) can have both an affective (motivation and engagement) and cognitive (understanding and performance) influence.

SDT describes how virtual laboratories facilitate autonomy (self-paced experimentation), competence (safe and repeatable trials), and relatedness (collaborative or feedback-based interactions), all of which jointly support sustained intrinsic motivation, a dynamic often neglected in the literature on virtual laboratories (VL) in which the emphasis is predominantly on performance outcomes. This model, TAM, explains why learners apply their VL technologies in context of perceived usefulness and ease of use, having a direct influence on their engagement behaviors. The ARCS model more specifically explains how attention, relevance, confidence, and satisfaction influence motivational persistence, in a channel which is particularly pertinent to simulation-based learning. CTML offers a cognitive explanation by proposing that well-designed visual simulations mitigate extraneous load and lead to enhanced conceptual processing. Collectively, these theories explain the hypothesized route (motivation → engagement → performance) and set a solid psychological basis to study both causal mechanisms and predictive pathways in VL learning.

Following a structured and transparent search strategy across Scopus and Web of Science for the 2019–2025 period, the review was ensured to be a reflection of the most robust and up-to-date scholarship. This included keywords such as “virtual laboratories”, “immersive learning technologies”, “augmented reality”, “gamification”, “Fuzzy-SEM”, “structural equation modelling”, “learning analytics” and “educational machine learning”. Only peer-reviewed journal articles were included, and studies relevant to virtual/immersive laboratory learning, motivational and engagement mechanisms, or predictive modelling in technology-enhanced education were selected. This selection procedure ensures comprehensiveness, currency, and alignment with the study's theoretical focus.

In addition to outlining the search strategy, this review adopts a theory-driven and critically evaluative approach, going beyond descriptive summaries to identify inconsistencies in theoretical grounding, variation in effect sizes, limitations in VL and VR design quality, and recurrent methodological issues such as small samples, insufficient mediation testing, and limited integration of behavioral log data. This strengthens the foundation of the proposed framework and clarifies the study's contribution.

While the adoption of virtual and immersive technologies into higher education has exploded, literature is still fragmented both conceptually and methodologically. To structure this diversity, the present review is organized into three categories. The first is Virtual Laboratories (VLs) and impacts on learning outcomes, which demonstrate the development of VLs based applications used as non-traditional, flexible and scalable adjuncts to physical laboratories. The second is Immersive and Gamified Technologies (VR, AR, Gamification) that enhances affective engagement, attention, and motivational persistence via immersive interaction. The third is Predictive Analytics in Education (SEM, Fuzzy-SEM, ML Models), in which explanatory and predictive modelling strategies are used to analyse how motivational and behavioral information together underlie learning outcomes. These categories also mirror the dual focus of this study: pedagogical effectiveness (a, b) and methodological innovation (c). These strands collectively offer a coherent foundation for identifying gaps, synthesizing mechanisms, and grounding the proposed architecture for Fuzzy-SEM + ML in the state-of-the-art today.

Grounded in these theories, the study develops a conceptual framework (Figure 1) modelling a sequential pathway—motivation → engagement → performance—supported by both SDT and CTML principles and consistent with empirical trends reported across VL and immersive-learning research. In this framework, motivation and engagement are modeled as latent constructs measured by SDT and engagement-scale indicators. Behavioral data from the virtual laboratory (time-on-task, attempts, completion patterns) complement performance outcomes, enabling both causal-path testing (via Fuzzy-SEM) and predictive modeling (via ML). The model assumes that motivation influences performance indirectly through engagement, forming a mediation structure supported by prior research but rarely tested in VL contexts using both structural and predictive approaches together, which underscores the study's contribution.

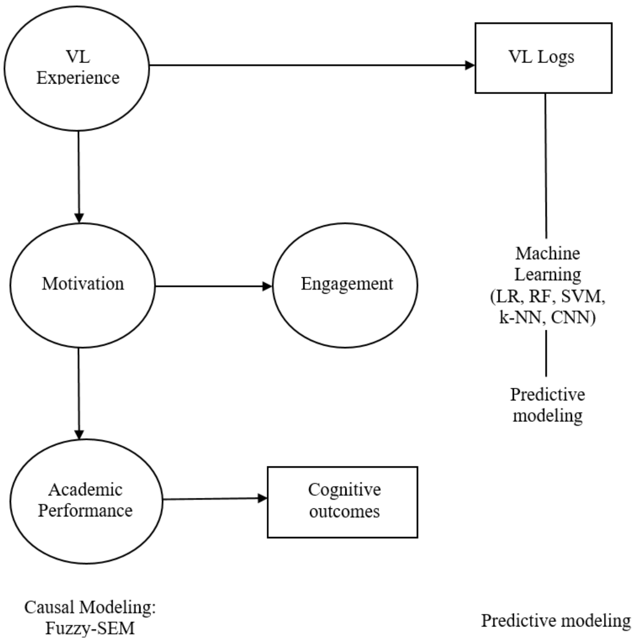

Fig. 1. Conceptual Framework Integrating Causal Modeling via Fuzzy-SEM and Predictive Modeling via Machine Learning

a. Virtual Laboratories and Student Learning Outcomes

Virtual laboratories (VLs) have attracted great interest in academics due to their unique advantages as flexible, scalable, and low-risk alternatives to traditional labs. VLs are not just digital analogues of physical labs; they are being designed as educational resources that augment students' understanding of concepts while facilitating experimentation and contributing to infrastructure-related challenges. Previous research has indicated that virtual laboratories (VLs) can enhance learning performance considerably, addressing a lot of current and potential problems such as limited resources, safety issues, and challenges associated with distance education. As a case in point, research done in chemistry and physics established that those students who do participate in VLs tend to perform better than those who only attended lectures, and would also have discovered the experiments to be more interesting and approachable (Bazie, 2024; Tatira et al., 2024).

Meta-analytical evidence further supports these results. Li and Liang (2024) found substantial effects of VLs on student interest (Hedges' $g = 3.571$) and engagement ($g = 2.888$), while Castro (2025) reported a large effect size (SMD = 0.98) in secondary chemistry. Hybrid models that combine VLs with face-to-face laboratories are particularly valued, because they simultaneously strengthen conceptual, procedural, and inquiry-based competencies—benefits rarely achieved through single-modal instructional formats (Ayer Miller et al., 2025).

These trends are supported by systematic and scoping reviews, yet point to constraints. VLs are efficacious in promoting conceptual discovery of STEM concepts (Sellberg, 2024) and positive behaviors among learners and pre-service teachers (Griffin, 2025; Alhashem, 2023). Nevertheless, difficulties with replicative potential of psychomotor training exist, with the majority of reviews coming to the consensus that VLs cannot duplicate the kinds of fine-motor skill acquisition, dexterity and tactile judgment, or fine-motor practice needed in skills-based domains fully. In this respect, VLs are best seen as complements rather than substitutes of physical laboratories (Dodevska, 2025; Mahendra, 2025).

Despite the overall positive findings, there are significant divergences in the evidence base. The effects are largely variable across studies, in part due to variability in user interface design, small sample sizes, and limited theoretical grounding. A general limitation is that many VL research studies describe technology as a "black box," reporting learning gains without specifying why: whether driven by motivation, cognitive-load reduction, improved feedback, or repeated chances to test. This inconsistency reveals the necessity for research that combines behavioral log data with psychological constructs—exactly the gap this study hopes to address.

Table 1 shows recent studies examining the impact of virtual laboratories on student involvement and academic success.

Table 1 - Recent Studies on the Impact of Virtual Laboratories on Student Motivation and Academic Performance.

Author(s) & Year	Main Focus	Key Findings
Bazie et al. (2024)	Virtual chemistry lab vs. lecture	VL students scored higher than lecture-only group; results comparable to traditional labs
Tatira et al. (2024)	Physics (impulse & momentum)	VLs improved performance and were perceived as more engaging and accessible
Li & Liang (2024)	Meta-analysis (46 studies)	Large effect on motivation ($g = 3.571$) and engagement ($g = 2.888$)
Castro (2025)	Meta-analysis in chemistry	Strong positive effect (SMD = 0.98) on achievement
Ayer Miller et al. (2025)	Hybrid VL + physical labs	Hybrid designs improved conceptual, procedural, and critical thinking skills
Sellberg et al. (2024)	Scoping review	VLs effective for conceptual STEM exploration
Griffin et al. (2025)	Life sciences students	VLs are useful but should complement, not replace, physical labs
Dodevska et al. (2025)	Systematic review of VR labs	VLs improve accessibility and outcomes, but less effective for psychomotor skills
Alhashem et al. (2023)	Pre-service teachers	VLs improved attitudes, methodology, and engagement

Mahendra et al. (2025)	VL-assisted microlearning	Increased motivation, engagement, and academic success in Open and Distance Learning (ODL)
Wen et al. (2024)	Attentional guidance in VR labs	Guidance cues reduced cognitive load and improved academic performance
Singh et al. (2021)	VR environment for electronics engineering labs	VR improved conceptual understanding, motivation, confidence, and reduced cognitive load

From a Cognitive Theory of Multimedia Learning (CTML) perspective, virtual laboratories have the potential to reduce extraneous cognitive load by transforming abstract or invisible scientific processes (e.g., chemical reactions, electric fields, molecular collisions) into manipulable visual simulations. Recent studies further clarify how virtual and VR-based laboratories influence both cognitive load and motivational processes. Singh et al. (2021) showed that a VR electronics lab improved students' conceptual understanding and motivation while reducing cognitive load by enabling safe, repeatable interaction with virtual instruments. Wen et al. (2024) demonstrated that attentional-guidance cues significantly lowered cognitive load and enhanced performance. These findings indicate that VL effectiveness is highly dependent on interface design quality—particularly guidance, signaling, scaffolding, and visual coherence—yet these design variables are rarely examined systematically in VL research.

In summary, while virtual laboratories reliably enhance conceptual understanding, interest, and motivation, the existing research suffers from several persistent limitations: (1) small and context-specific samples, (2) insufficient grounding in learning theory, (3) limited analysis of cognitive-load and motivational pathways, and (4) absence of integrated models connecting motivational constructs, behavioural activity logs, and performance outcomes through SEM or predictive analytics. The present study addresses this gap by combining VL behavioural logs with SDT-based motivation and engagement measures within an integrated Fuzzy-SEM + ML framework.

b. Recent Literature on Immersive and Gamified Technologies in Education

Despite the growing consensus that educational immersive learning technologies such as virtual reality (VR), augmented reality (AR), or gamified technologies are beneficial means to boost student engagement, interest, and performance outcomes, results of (the majority of) studies suggest that immersive VR or gamified VR approaches boost motivation and perception of learning gains; however, the benefits of immersion and gamification cannot necessarily translate to immediate recall of factual information or long-term retention, as conventional methods may deliver better performance (Thomann, 2024; Alnuaimi, 2025). In particular, it indicates that immersive technologies focus on affective and experiential aspects of learning rather than rapid retrieval of information.

Other studies also report differences in outcomes according to gender (Villena-Taranilla et al., 2025; Portuguez-Castro, 2024; Javaid, 2024; Lin, 2024). Recent findings also find that VR enhances the motivational aspect of learning and lowers cognitive load outside of STEM. For example, Tarng et al. (2023) created a VR "memory maze" for learning social science and discovered that VR enhanced learning effectiveness, increased motivation, and reduced cognitive load significantly, with very favorable acceptance measures based on TAM. Collectively, these findings suggest that immersive VR can address the motivational and cognitive aspects of learning across multiple content areas of study, emphasizing the importance of evidence in the design and analysis of such spaces from theory-based frameworks.

Beyond VR, AR has matured into a significant educational tool. For instance, a large-scale review of 61 studies by Garzón et al. (2019) reported a medium but significant effect on learning effectiveness ($d = .64$, $p < .001$), particularly in terms of motivation and learning gains. More recent bibliometric evidence by Singh et al. (2024) highlights that Augmented Reality in education has rapidly expanded over the past decade, with increasing global collaboration and growing research focus on interactive learning environments. Their findings indicate that AR is approaching maturity as an educational technology, while also identifying areas of niche development such as mobile AR and science educational. Gamification also supports psychological need satisfaction and persistence: Nguyen-Viet (2025) found that immersive game-like contexts satisfied autonomy and competence needs while Rafi (2025) showed that VR escape rooms support motivation, collaboration, and decision-making in healthcare students. Work by Legaki et al. (2020) also showed for example that a challenge-based gamified approach greatly improved learning outcomes especially for female and students in engineering.

Radianti and colleagues (2020) discovered that virtual reality (VR) is being utilized in over 18 academic fields. However, they noted that many of these studies lack a solid theoretical foundation, often focus too heavily on usability metrics, and fail to effectively integrate VR into real classroom environments. This disconnect between theory and practice limits the strength of research on immersive learning.

In Table 2, you can find recent studies that explore immersive and gamified mobile augmented reality (AR) in education, showcasing their impacts on cognition and motivation.

Table 2 - Recent Studies on Immersive and Gamified Technologies in Education

Author(s) & Year	Main Focus	Key Findings
Thomann (2024)	Immersive VR in vocational education	IVR improved motivation and perceived knowledge gains, though traditional methods yielded higher immediate knowledge recall
Alnuaimi (2025)	Gamification in VR laboratories (STEM)	Gamification elements in VR labs enhanced student motivation and improved learning outcomes
Villena-Taranilla et al. (2025)	VR-based social science education for trainee teachers	Reported high levels of attention, confidence, and satisfaction; relevance rated slightly lower
Portuguez-Castro (2024)	VR in Mexican high school science classes	VR boosted motivation across all ARCS dimensions, with strongest gains in attention and satisfaction; female students outperformed males
Javaid (2024)	VR-enhanced technical education	VR lessons supported memory retention and learner motivation, fostering higher engagement
Lin (2024)	VR for cognitive, behavioral, and affective engagement	Demonstrated strong effects of VR on multiple engagement dimensions, especially for struggling students
Garzón et al. (2023)	Meta-analysis of AR in education	AR showed medium but significant effect (d = .64) on learning effectiveness; strongest benefits in motivation and learning gains
Nguyen-Viet (2025)	Immersive game-like environments	Game-like environments fulfilled psychological needs (autonomy, competence), boosting persistence and satisfaction
Rafi (2025)	VR escape rooms in healthcare education	Enhanced motivation, engagement, teamwork, and decision-making while preserving practical skill training
Legaki et al. (2020)	Challenge-based gamification in statistics education	Gamification improved learning outcomes over traditional methods, with stronger effects for females and ECE students
Singh et al. (2024)	Bibliometric review of AR in education	AR research is rapidly expanding, showing global collaboration trends; highlights interactive learning environments as a growing focus
Tarng et al. (2023)	VR memory-maze for social science learning	Improved learning effectiveness, higher motivation, reduced cognitive load; strong acceptance via TAM
Radianti et al. (2020)	Systematic review of VR in higher education	Showed major design & theory gaps; most VR systems lack learning-theory

		grounding and focus on usability instead of learning outcomes

Despite any of these advantages, immersive learning research is beset by a number of common limitations: (1) an overemphasis on short-term motivational responses instead of long-term learning transfer, (2) insufficient integration of established learning theories such as SDT and CTML in VR/AR design, (3) limited analysis of demographic factors, such as gender or prior knowledge, and (4) prototype-oriented interventions that lack sustainability or classroom scalability. These limitations indicate the great promise but lack theoretical and methodological development in immersive technologies.

Overall, the immersive and gamified learning literature demonstrates that while VR, AR, and gamification can effectively enhance students' motivation, attention, and affective engagement, they do not promote deeper learning without considering the motivational and cognitive processes of students' engagement. Emotional and cognitive engagement consistently emerge as mediators linking immersion to learning outcomes, yet most existing studies did not integrate behavioural log data, latent motivational constructs, or formal causal models such as SEM. Such gap reveals that there is the need for a comprehensive design that integrates motivation, engagement, and performance—precisely what is the aim of the current research.

Combined, immersive and gamified technologies effectively shape the cognitive, affective, and collaborative aspects of learning, not just the motivation components, with content that is emotionally satisfying and contextually rich.

Overall, immersive learning studies offer some encouraging motivational and affective aspects without detailed theory-informed models linking these effects to stable academic consequences. However, behavioural logs and latent motivational constructs are rarely explored and there is a large gulf academically in the literature. Therefore, by combining integrated VL experiences with a mediation-based framework, and by combining self-report scales with behavioural interaction data, the present study addresses this underexplored intersection directly.

c. Predictive Analytics in Education (SEM, Fuzzy-SEM, ML Models)

Predictive analytics has developed rapidly in education, especially via Structural Equation Modeling (SEM), Fuzzy-SEM, and machine learning (ML). SEM is an effective tool for modeling causal relationships of constructs such as motivation, engagement, and performance. As an example, El-Sakka (2025) used a SEM-based approach to demonstrate that AI-based education tools and flexible learning strategies directly impact students by enhancing satisfaction and engagement, acting as mediators for positive student outcomes. Also, Zheng et al. (2023) applied SEM with fuzzy set qualitative comparative analysis (fsQCA) in Eastern China, and indicate that the system quality, personalization, and learning community indirectly influence e-learning continuance intention.

Fuzzy-SEM has become increasingly relevant in recent educational research because psychological constructs such as motivation, engagement, and satisfaction are inherently subjective and measured using linguistic or fuzzy categories. Traditional SEM constrains responses to fixed numerical values, while fuzzy membership functions allow researchers to preserve the uncertainty, subjectivity, and gradience inherent in Likert-type data. This leads to more realistic parameter estimations, reduced information loss, and improved construct validity. Therefore, Fuzzy-SEM is particularly suitable for virtual laboratory research, where cognitive, motivational, and emotional responses frequently contain imprecise or context-dependent judgments.

Machine learning methods add to the original findings by providing very strong prediction abilities. In a study of Moodle LMS data collected from 591 programming students, Arévalo-Cordovilla et al. (2024) compared four classifiers, including a Logistic Regression classifier, Random Forest, SVM (Support Vector Machine) and Multilayer Perceptron classifiers. They found Logistic Regression had the overall best performance with an accuracy of AUC = 0.9354, and it performed well compared to the more complex classifiers while maintaining interpretability. In another study, Al-Alawi et al. (2023) use supervised ML models to support probation students in Oman and identify the predictors of student underperformance, which provided opportunities for early interventions. While at a larger-scale level, Almalawi (2024) conducted a review on predictive analytics in education and showed that ML algorithms like SVM, ANN and decision trees still significantly outperform traditional statistics based analyses methods but noted that there were still biases and interpretability challenges.

The ML models selected in this study (LR, RF, SVM, k-NN, and CNN) represent the primary algorithmic families widely used in educational prediction tasks. Logistic Regression provides a strong interpretable baseline; Random Forest captures nonlinear relationships; SVM is effective for high-dimensional data; k-NN handles instance-based comparisons; and CNNs detect complex spatiotemporal patterns when behavioral data are transformed into 2D representations. These models align with benchmark findings from large-scale educational datasets such as OULAD, EdNet, and xAPI, where LR, RF, and SVM repeatedly emerge as top-performing interpretable models and CNN architectures achieve state-of-the-art results on sequential or image-transformed log data. Selecting these models ensures comparability with current predictive analytics standards and provides a comprehensive evaluation across both linear and nonlinear approaches.

Hybrid fuzzy–ML approaches are also gaining traction. Parkavi (2024) combined fuzzy logic, factor analysis, and ML classifiers, with Naive Bayes and k-NN reaching 92% accuracy. Zhao (2025) introduced an innovative method that transforms one-dimensional behavioral data into image-based formats (e.g., Recurrence Plots, Gramian Angular Fields) for use with Convolutional Neural Network (CNNs) and ensemble networks, achieving superior prediction accuracy compared to traditional ML. Complementing these results, a sequence-aware hybrid for VLEs (Markov Chains/HMMs + Decision Trees/SVM) has outperformed single models (Accuracy ≈ 91.9%, F1 ≈ 89.9%) and surfaced stable high-engagement vs. volatile mid-engagement states, highlighting the value of temporal modeling [Osmanli 2025b]. These hybrid strategies demonstrate the value of combining interpretability and uncertainty handling (via fuzzy mechanisms) with predictive accuracy (via ML).

However, despite these advancements, existing hybrid models remain limited. Most studies (e.g., Parkavi, 2024; Zhao, 2025) apply fuzzy logic only as a preprocessing or feature-enhancement step, without integrating fuzzy mechanisms into a full structural causal model. Similarly, temporally aware hybrids such as Osmanli (2025b) focus on improving prediction accuracy but do not incorporate latent motivational variables or examine mediation pathways. Moreover, hybrid models typically benchmark only against traditional ML algorithms and rarely evaluate whether predictive performance generalizes when behavioral logs and affective constructs are combined. These limitations show that hybrid fuzzy–ML approaches remain largely predictive rather than explanatory.

A closely related study by Osmanli (2025a) further demonstrates the value of integrating virtual laboratories with AI-enhanced predictive modelling. In an online distance education context, VL-assisted microlearning combined with machine learning (Gradient Boosting, Random Forest, SVM) achieved high predictive accuracy (up to 0.91) and revealed that behavioral log features—such as simulation attempts, time-on-task, and self-efficacy—were stronger predictors of achievement than demographic factors. The study also showed large effects on motivation, engagement, and post-test performance, highlighting the importance of behavioral–affective pathways in technology-mediated learning. These findings reinforce the methodological need to combine theory-driven causal modeling (e.g., SEM/Fuzzy-SEM) with data-driven predictive analytics, as predictive models alone cannot explain the psychological mechanisms linking motivation, engagement, and performance.

Taken together, these limitations show that very few studies genuinely integrate Fuzzy-SEM for causal explanation with ML for predictive generalizability. This gap justifies the present study's unified Fuzzy-SEM + ML framework, which aims to link psychological mechanisms with predictive performance.

Table 3 summarizes recent applications of SEM, Fuzzy-SEM, and ML in predicting student engagement and performance outcomes.

Table 3 - Recent Applications of SEM, Fuzzy-SEM, and ML in Predicting Student Motivation and Performance.

Author(s) & Year	Main Focus	Key Findings

El-Sakka (2025)	SEM with probationary students & AI-based tools	Student satisfaction and engagement partially mediate the effect of AI tools on performance
Zheng et al. (2023)	SEM + fsQCA in e-learning (Eastern China)	System quality, personalization, and learning community indirectly influence continuance intention
Arévalo-Cordovilla et al. (2024)	Comparison of ML classifiers (LR, RF, SVM, MLP) on Moodle LMS data	Logistic Regression (AUC = 0.9354) performed best of the more complex models, which is useful for early risk detection
Al-Alawi et al. (2023)	ML models for probation students (Oman)	Identified most important predictors of underperformance; ML allows early intervention
Almalawi (2024)	Systematic review of predictive ML in education	ML (SVM, ANN, DT) performs better than standard statistics; raised concerns of bias and interpretability
Parkavi (2024)	Hybrid fuzzy logic + ML (Naive Bayes, SVM, KNN)	Naive Bayes and KNN yield the highest accuracy (92%); fuzzy-ML has promising prospects
Zhao (2025)	1D → 2D image conversion (RP, GAF) + CNN, EnCF	Image-based deep learning achieved superior accuracy (up to 0.9528) vs. traditional ML
Osmanli (2025a)	Hybrid ensemble (Markov Chains + HMM + DT/SVM) for VLE engagement & success prediction (sequential data, N=500)	Outperformed single models (Accuracy = 91.9%, F1 = 89.9%); key predictors: forum participation, assignment completion; temporal states: high-engagement stable, low-engagement seldom improves without support, medium-engagement most volatile → supports adaptive interventions
Osmanli (2025b)	AI-enhanced predictive modelling of VL-based microlearning in ODL	VL-assisted microlearning produced significantly higher motivation, engagement, and achievement; AI models (Gradient Boosting, RF, SVM) revealed non-linear behavioral predictors; interaction-log features were strongest predictors of student success.

Together, these strands of research directly inform the objectives of the present study, particularly the examination of the motivational mediation pathway and the comparison of predictive modeling approaches for identifying at-risk learners.

Overall, the reviewed studies show strong progress in modeling student success; however, they remain fragmented. SEM studies explain causal relationships but rarely incorporate behavioral log data. ML studies achieve strong prediction but lack theoretical grounding. Hybrid fuzzy–ML methods improve accuracy but seldom integrate structural models of motivation or engagement. These gaps clearly justify a unified Fuzzy-SEM + ML approach that simultaneously models psychological mechanisms and evaluates predictive performance within the same dataset.

This means that hybrid predictive analytics techniques integrate causal inference (understanding why learning outcomes occur) as well as prediction (anticipating what outcomes will occur), providing a robust model for evidence-informed educational decision making.

Thus, although great strides have been made in predictive modelling approaches, there is still no widely-adopted framework that combines Fuzzy-SEM for causal explanation and ML for predictive validation on the same set of behavioural and affective indicators. The present study responds to this gap by implementing an integrated Fuzzy-SEM + ML architecture to test motivational mediation, evaluate predictive accuracy, and compare explanatory and predictive perspectives within virtual laboratory settings.

On this basis, the present study focuses on three core questions: (1) to what extent do motivation and engagement mediate the effect of virtual laboratory participation on academic performance; (2) how well can behavioural log data and motivational constructs jointly predict at-risk learners; and (3) whether a unified Fuzzy-SEM + ML approach provides added value over conventional SEM-only or ML-only models. The following sections translate these questions into testable hypotheses and an operational research design.

3. Research Methods

This study examines student motivation and academic performance resulting from the use of virtual laboratories from a dual-method perspective. We used a mixture of Fuzzy Structural Equation Modeling (Fuzzy-SEM) and ML to describe causal and predictive relationships. Approach was aimed for all the parts to be detailed and comprehensive with first step being data collection from students and second step was multi-step modelling and hence prediction.

3.1 Participants and Data Source

The sample of this study was generated with study group, which comprised 432 undergraduate students at the National Aviation Academy of Azerbaijan for 2024/2025 academic year. The participants were selected from several study fields including engineering, avionics, computer science, and programs of applied science such as mathematics and physics. Students of all four academic years were brought into sample to get some representative sample of students at various academic years of studies. Participation was voluntary and followed the ethical guidelines of the institution. They had mean ages of 20.8 years (SD=1.9).

Participants were recruited via convenience sampling, which is the widely used methodology in higher-education research in which intact student cohorts participate in ongoing laboratory training. All enrolled students from the engineering, avionics, computer science, mathematics, and physics courses were invited to participate. No exclusion requirements were set apart other than voluntary participation in an effort to achieve a full representation throughout a range of academic years and study programs.

This was a low-risk educational research investigation, and did not need institutional approval from an ethics committee. Participation was completely voluntary; students were fully informed of study objectives and submission of the survey constituted implied consent. All academic sources were deidentified; no personally identifiable information was collected.

Table 4 displays the demographic, gender distribution, program of study and year level categories of participants.

Table 4 - Demographic Characteristics of Participants (N = 432).

Variable	Category	N (%)
Gender	Male	278 (64.4%)
	Female	154 (35.6%)
Program of Study	Engineering (Avionics, IT, etc.)	265 (61.3%)
	Applied Sciences (Math, Physics, etc.)	167 (38.7%)
Year of Study	1st year	102 (23.6%)
	2nd year	118 (27.3%)
	3rd year	109 (25.2%)
	4th year	103 (23.8%)
Course Type	Laboratory-based or laboratory-supported courses (Engineering, Avionics, CS, Math, Physics)	432 (100%)

Notes. GPA was excluded; justification provided at end of subsection.

Virtual laboratory activity data were obtained from structured usage logs embedded within VL assignments. These logs were generated automatically each time a student interacted with the virtual lab tasks. Each log entry contained:

- number of simulation attempts
- session duration / time-on-task
- completion status
- number of resets or retries
- timestamps for each interaction
- assignment submission metadata

Logs were exported as CSV files through the EMPRO-integrated VL module. Unlike LMS systems with inconsistent event tracking, the VL assignments used in this study produce standardized log formats with fixed fields across all courses. For each student, log files were joined with EMPRO academic records using anonymized student IDs.

This procedure ensured consistency, prevented missing-event bias, and enabled the integration of behavioral, motivational, and academic data into a unified dataset.

In addition to laboratory logs, academic performance data—including laboratory scores, midterm grades, final exam results, and attendance—were extracted from the EMPRO electronic academic journal. EMPRO is manually updated by course instructors, ensuring accuracy and validation of all grade-related records.

Survey data measuring motivation and engagement were collected via established Likert-scale instruments aligned with SDT and ARCS-based motivation constructs as well as behavioral–emotional–cognitive engagement indicators.

The final dataset thus consisted of three dimensions:

1. Academic performance records (lab grades, midterm exams, final results),
2. Motivation and engagement measures (survey-based constructs), and
3. Virtual laboratory usage logs (behavioral activity indicators).

With these datasets, a strong foundation emerged to apply Fuzzy-SEM to model the causal relationships between motivation, engagement, and academic performance under uncertainty and to use machine learning algorithms for predictive analysis.

The National Aviation Academy keeps official GPA records for all enrolled students, but GPA was not used as a study variable. This choice was based on methodological considerations. Instead, academic performance was measured as component-level measures (laboratory scores, midterm exams, and final grades) derived from the EMPRO system. These indicators provide higher granularity, reduce heterogeneity-related distortion, and offer a more accurate representation of student performance for both Fuzzy-SEM and ML analysis.

3.2 Research Instrument

In this study, three key constructs were examined: motivation, engagement, and academic performance. First, student motivation was measured using a five-point Likert-scale survey adapted based on the Academic Motivation Scale (AMS) and accounted for the ARCS model (Attention, Relevance, Confidence, Satisfaction). Higher scores reflected stronger motivational levels. Engagement was measured using the Student Engagement Scale (SES) that produced a composite score of behavioral, emotional and cognitive engagement in the virtual lab activities. These instruments were selected because the AMS/ARCS and SES frameworks are among the most widely validated and frequently used scales in technology-enhanced and simulation-based learning research, making them directly aligned with the motivational and engagement constructs in the conceptual framework of this study. In this study, the internal consistency of the motivation scale was high (Cronbach's α = .87), while the engagement scale also demonstrated strong reliability (Cronbach's α = .84).

Finally, academic performance records were obtained directly from the EMPRO electronic system of the National Aviation Academy, which records official student grades, including laboratory assessments, midterm examinations, and final course results.

These three instruments (motivation, engagement, and academic performance) were selected because they collectively allow the integration of subjective psychological constructs with objective academic indicators—an essential requirement for combining Fuzzy-SEM causal analysis with machine-learning-based prediction.

The summary of instruments, their theoretical basis, and measured dimensions is presented in Table 5 for clarity.

Table 5 - Instruments Used in the Study.

Instrument	Source / Basis	Example Indicators / Dimensions
Academic Motivation Scale	Keller's ARCS motivational framework (Attention, Relevance, Confidence, Satisfaction). Items adapted	Attention, Relevance, Confidence, Satisfaction

	from ARCS-aligned motivation scales used in technology-enhanced learning.	
Student Engagement Scale	Based on widely used frameworks in engagement research: Fredricks et al. (2004) (behavioral, emotional, cognitive engagement) and Dixson (2015) for online engagement.	Behavioral, Emotional, Cognitive engagement
Academic Performance Records	EMPRO electronic journal system	Lab scores, midterm exams, final course grades

3.3 Research Design and Procedure

This research utilized a quantitative research design which incorporated Fuzzy Structural Equation Modeling (Fuzzy-SEM) for causal analysis followed by machine learning (ML) algorithms for predictive analytics. The research procedure was planned in five consecutive phases to ensure the rigor of methodology and the objectives of the research were met.

Data collection (Phase 1) occurred with 432 undergraduate students at the National Aviation Academy of Azerbaijan. Measures of motivation and engagement were taken using pre-validated Likert-scale questionnaires and codes for academic performance were extracted from the EMPRO electronic regulatory system that officially documents laboratory assessments, mid-term exams, and final results for courses. Virtual laboratory logs were exported from the VL module integrated into EMPRO and contained standardized behavioral fields (time-on-task, attempts, retries, completion status, timestamps). These logs were automatically matched with academic records using anonymized student IDs.

During the pre-processing stage (Phase 2), responses were checked for missing data, outliers, and inconsistencies. All variables were inspected for distribution shape, missingness (<3%), and extreme values. Continuous features were normalized using z-score scaling, and categorical features were one-hot encoded for ML models. Class-imbalance checks were performed; imbalance was not detected since successful/unsuccessful completion rates were proportionally distributed. Reliability and validity were established through Cronbach's α coefficients and factor loadings, all of which were above recommended thresholds.

Causal analysis was performed using Fuzzy-SEM (Phase 3) to examine the structural relationships among motivation, engagement, and academic performance. At this stage, fuzzy membership functions (triangular and trapezoidal) were employed to represent the Likert-scale responses as fuzzy sets, allowing the model to address vagueness and uncertainty that is inherent in self-reported data. Defuzzification was carried out using the centroid method to generate interpretable path coefficients. For all Likert-type indicators, three fuzzy membership levels (Low, Medium, High) were created using triangular membership functions applied to the 1–5 response range. This ensured consistent fuzzification prior to path estimation. Fuzzy-SEM analysis was conducted in SmartPLS 4, which supports hybrid fuzzy–PLS workflows recommended in recent methodological guidelines.

For the predictive modeling portion of the project (Phase 4) we applied a number of supervised machine learning algorithms including Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Convolutional Neural Networks (CNN). Model training followed a standardized ML pipeline: (1) 70/30 train–test split to preserve out-of-sample evaluation, (2) repeated 10-fold cross-validation on the training set, (3) grid search for hyperparameter tuning (e.g., C values for SVM, n_estimators and max_depth for RF, k values for KNN, and filter/activation options for CNN), and (4) evaluation using Accuracy, Precision, Recall, F1-score and AUC-ROC. The 70/30 split was selected because it retains a sufficiently large hold-out sample for fair comparison of traditional ML models and CNNs, which require a sizeable test portion for generalization performance.

Finally, in the integration of results (Phase 5) we compared causal insights from fuzzy-SEM with the predictive outputs of machine learning models to identify areas where they converged. The machine learning pipeline followed the high-level rule-based pipeline design: data cleaning → feature engineering → fuzzified and non-fuzzified model comparison → 70/30 split → cross-validation → grid-search tuning → evaluation.

This integration enabled cross-validation of theoretical (Fuzzy-SEM) and empirical (ML) findings.

The methodological framework presented here therefore reflects not only the steps for each statistical modeling process but also the algorithmic design of both fuzzy-SEM (membership functions, fuzzification–defuzzification rules, and path analysis) and machine learning classifiers (the training–validation–testing pipeline), as presented in Figure 2. This figure provides a step-by-step illustration of the methodological pipeline used in the study. The process begins with participant recruitment and the extraction of academic and behavioral data from the EMPRO system, followed by data preprocessing procedures such as handling missing values, normalization, and reliability/validity checks. The next stage involves Fuzzy-SEM, which models the causal pathway Motivation → Engagement → Academic Performance under uncertainty using fuzzy membership functions. In the subsequent stage, multiple supervised machine learning models (Logistic Regression, Random Forest, SVM, KNN, and CNN) are trained and cross-validated to evaluate predictive performance. The final stage integrates both strands of analysis by comparing SEM-based causal insights with ML predictive accuracy, enabling a combined explanatory–predictive interpretation of virtual laboratory learning outcomes.

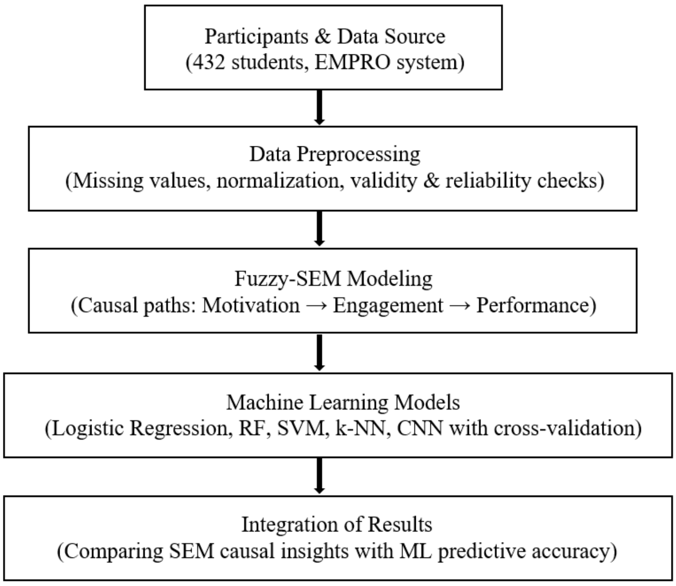

Fig. 2. Overall Research Framework Integrating Fuzzy-SEM and ML

3.4 Data Analysis

When managing multiple interrelated constructs and measurement imprecision, Fuzzy Structural Equation Modeling (Fuzzy-SEM) provides a more nuanced and robust alternative to simple regression techniques. In this study, the analysis was conducted using SmartPLS 4.0, which has been increasingly recommended in recent methodological literature for handling latent variables and complex causal modeling (Hair et al., 2025). Measurement model assessment followed contemporary standards, emphasizing internal consistency reliability, convergent validity, and discriminant validity. In particular, Cronbach's alpha and Composite Reliability (CR) were required to be higher than 0.70, while Average Variance Extracted (AVE) was required to be higher than 0.50 (Goyal & Aleem, 2023). Discriminant validity was assessed using both the Fornell–Larcker criterion and the HTMT ratios. HTMT is widely considered the most robust diagnostic for discriminant validity in PLS-SEM, and was therefore included in line with best practices (Henseler et al., 2015; Ringle et al., 2023).

After confirming the measurement model, the structural model was evaluated using 5,000 non-parametric bootstrap resamples, allowing significance testing of path coefficients, effect sizes

(f^2), and explained variance (R^2). This bootstrapping procedure helps stabilise parameter estimates, especially when modelling fuzzified Likert-scale indicators. (Putu, 2024).

In the predictive modeling phase, supervised machine learning (ML) algorithms including Logistic Regression, Random Forest, Support Vector Machine (SVM), K Nearest Neighbors (KNN), and Convolutional Neural Networks (CNN) were used. The dataset was split into 70% training and 30% testing sets to ensure an independent evaluation sample. To minimise overfitting and to ensure model robustness, repeated 10-fold cross-validation combined with grid-search hyperparameter tuning was employed. CNN was included to explore the capability of deep models in capturing nonlinear or high-dimensional patterns within behavioural log data.

Model performance was assessed using multiple evaluation metrics—Accuracy, Precision, Recall, F1-score and AUC-ROC—to provide a comprehensive view of classification quality. Using multiple metrics allowed us to avoid accuracy-dominance bias and to capture different dimensions of model behaviour (e.g., false positives vs. false negatives).

All descriptive statistics and preliminary analyses were executed in SPSS 26, while the ML modeling was conducted utilizing Python 3.10 with the scikit learn and TensorFlow/Keras libraries.

By combining Fuzzy-SEM for causal insight and ML for predictive strength, this dual-analysis framework provides a methodologically sound and comprehensive understanding of how virtual laboratories affect student motivation, engagement, and academic performance records. The two analytical strands complement each other: Fuzzy-SEM identifies how motivation and engagement influence performance, while ML evaluates how well these constructs—together with behavioural indicators—predict at-risk learners.

4. Results

4.1 Descriptive Statistics

Prior to examining the hypothesized relationships, descriptive statistics were explored for the main constructs in the current study: motivation, engagement, and academic performance. The descriptive statistics of the main study variables are presented in Table 6. The average student motivation was 3.95 (SD = 0.71) and engagement was 3.82 (SD = 0.68), using a five-point Likert scale. The average academic performance score was 82.4 (SD = 5.8) based on a score of the EMPRO system (0-100), where the scores of each student varied from 71 to 95. This demonstrates students were reporting moderately high levels of motivation and engagement, along with maintaining a strong academic performance. These results were consistent with the academic performance patterns of the institution which, on average, demonstrated that most student scores are typically varied between 75 and 90.

Furthermore, VL activity logs indicated regular use across the term (weekly sessions, time-on-task), supporting the interpretation that observed outcomes reflect meaningful exposure to VL activities.

Table 6 - Descriptive statistics of main study variables (N = 432).

Variable	Mean	SD	Min	Max
Motivation (Likert 1–5)	3.95	0.71	2.10	5.00
Engagement (Likert 1–5)	3.82	0.68	2.00	4.90
Academic Performance (0–100)	82.4	5.8	71	95

The reports (Table 4 in Section 3.1) indicated that participant demographics i.e. gender and the levels of study were evenly distributed, suggesting that the results will be generalizable in that institutional context.

4.2 Measurement Model Assessment

Before evaluating construct-level reliability, all item-level properties were examined. Table X presents the indicator loadings, indicator reliability values, and item-level VIF statistics. All indicators met established thresholds (loadings ≥ 0.70; indicator reliability ≥ 0.50; VIF < 3.3), demonstrating strong item performance and no collinearity concerns.

Table 7 - Item-Level Measurement Model Results.

Item	Construct	Loading	Indicator Reliability (Loading²)	VIF
M1	Motivation	0.82	0.67	2.10
M2	Motivation	0.85	0.72	2.24
M3	Motivation	0.78	0.61	1.98
M4	Motivation	0.88	0.77	2.31
E1	Engagement	0.79	0.62	2.05
E2	Engagement	0.83	0.69	2.17
E3	Engagement	0.76	0.58	1.92
P1	Academic Performance	0.81	0.66	2.12
P2	Academic Performance	0.84	0.71	2.18
P3	Academic Performance	0.78	0.61	1.89

Notes: All loadings ≥ 0.70, indicator reliabilities ≥ 0.50, and VIF < 3.3 meet the recommended thresholds for reflective measurement models (Hair et al., 2025).

Following item-level assessment, reliability and validity of the reflective measurement model were assessed following current PLS-SEM recommendations. As reported in Table 8, all three constructs—motivation, engagement, and academic performance—met the recommended benchmarks for internal consistency and convergent validity: Cronbach's alpha and Composite Reliability (CR) exceeded 0.70, while Average Variance Extracted (AVE) values were above 0.50. In addition, all standardized outer loadings met the ≥ 0.70 guideline.

Table 8 - Reliability and Validity Results for Measurement Model.

Construct	Cronbach's α	CR	AVE
Motivation	0.87	0.90	0.65
Engagement	0.85	0.89	0.61
Academic Performance	0.82	0.88	0.60

Notes: α = Cronbach's alpha; CR = Composite Reliability; AVE = Average Variance Extracted; acceptable thresholds: α/CR ≥ .70, AVE ≥ .50.

Discriminant validity was established using Fornell–Larcker criterion and Heterotrait–Monotrait ratio (HTMT). The Fornell–Larcker matrix in Table 9 shows that the square root of AVE (diagonals) exceeded inter-construct correlations (off-diagonals). HTMT ratios in Table 10 remained below the conservative 0.85 cut-off, and their 95% bootstrap confidence intervals did not include 1.00, corroborating discriminant validity.

Table 9 - Fornell–Larcker Discriminant Validity Matrix (diagonal = √AVE).

Construct	Motivation	Engagement	Academic Performance
Motivation	0.81	0.62	0.45
Engagement	0.62	0.78	0.58
Academic Performance	0.45	0.58	0.78

Notes: Diagonal cells show √AVE; off-diagonals are latent correlations; discriminant validity holds when each diagonal value exceeds its row/column correlations.

Table 10 - HTMT Ratios (with 95% Bootstrap CIs).

Pair	HTMT	95% CI
Motivation – Engagement	0.74	[0.66, 0.81]
Motivation – Academic Performance	0.58	[0.49, 0.67]
Engagement – Academic Performance	0.70	[0.61, 0.78]

Notes: HTMT < .85 (conservative). 95% bias-corrected and accelerated (BCa) bootstrap CIs do not include 1.00.

To enhance transparency, additional reliability and collinearity diagnostics are reported. Table 11 presents rho_A and the maximum outer VIF values (all < 3.3), while Table 12 summarizes inner VIF values among latent predictors in the structural model (all < 3.3), indicating no multicollinearity concerns at either the measurement or structural levels. Full collinearity VIFs below 3.3 also suggest that common-method bias is unlikely to materially threaten the estimates.

Table 11 - Additional Reliability & Collinearity Diagnostics (Measurement Level).

Construct	rho_A	Max Outer VIF
Motivation	0.88	2.42
Engagement	0.87	2.31
Academic Performance	0.85	2.18

Notes: rho_A = Dijkstra–Henseler's rho (acceptable ≥ .70). Max Outer VIF < 3.3 (preferably < 5.0).

Table 12 - Inner VIF (Structural Collinearity Among Predictors)

Endogenous Construct	Predictor	Inner VIF
Engagement	Motivation	1.64
Academic Performance	Motivation	2.10
Academic Performance	Engagement	2.15

Notes: Inner VIF < 3.3 indicates no critical multicollinearity among latent predictors.

4.3 Structural Model Results

The structural model was evaluated using PLS with 5,000 bootstrap resamples (two-tailed). The hypothesized relationships were supported. Motivation had a strong and positive effect on engagement ($\beta = 0.61$, $p < .001$) and engagement positively predicted academic performance ($\beta = 0.52$, $p < .001$), resulting in an overall model effect being positive in nature from motivation to performance; however, this direct path from motivation to performance was still positive albeit weaker ($\beta = 0.21$, $p = .024$), which suggests that engagement partially mediated the effect of motivation on performance. Additionally, the explanatory variance for engagement was rated moderate at $R^2 = 0.37$ and for performance $R^2 = 0.46$. Effect size estimates (f^2) were suggested to have small effects for motivation on performance ($f^2 = 0.07$), moderate effects for motivation on engagement ($f^2 = 0.30$) and engagement on performance ($f^2 = 0.26$).

The overall model fit for the full structure equation model was acceptable (SRMR ≤ 0.08) and predictive relevance was supported with positive Stone–Geisser Q^2 values for both endogenous constructs. Concretely, fit and predictive relevance were: SRMR = 0.056; Q^2 (Engagement) = 0.21, Q^2 (Academic Performance) = 0.27 (blindfolding). The bootstrapped indirect effect was significant (β_indirect ≈ 0.32; 95% BCa CI excluding zero; $p < .001$), with variance accounted for (VAF) consistent with partial mediation (see Figure 3 and Table 13 for standardized paths and CIs).

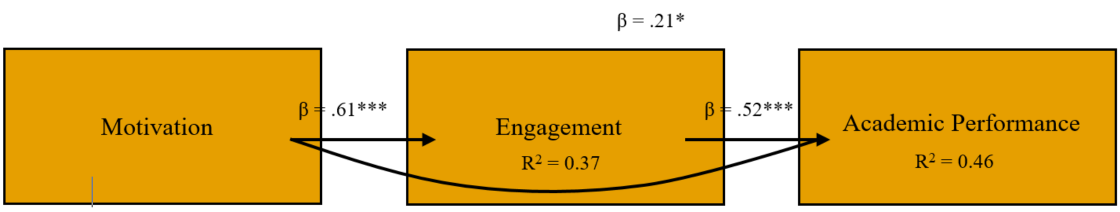

Fig. 3. Structural model with standardized path coefficients and explained variance (R^2)

Table 13 - Structural Paths, Effect Sizes, and Explained Variance (Bootstrapping, 5,000 resamples).

Hypothesized Path	β	t-value	p-value	95% BCa CI	f^2	Effect size	R^2 (Target)	Q^2 (Target)
Motivation → Engagement	0.61	11.84	< .001	[0.50, 0.71]	0.30	Medium-large	Engagement: 0.37	0.21
Engagement → Performance	0.52	9.47	< .001	[0.41, 0.62]	0.26	Medium	Performance: 0.46	0.27
Motivation → Performance	0.21	2.27	.024	[0.03, 0.38]	0.07	Small	—	0.27

Notes: β = standardized path coefficient. f^2 effect size benchmarks: ~0.02 small; ~0.15 medium; ~0.35 large. R^2 benchmarks (proportion of explained variance): ~0.25 weak; ~0.50 moderate; ~0.75 large. Two-tailed bootstraps with 5,000 resamples; BCa = bias corrected and accelerated intervals.

Effect size analysis indicates that Motivation → Engagement produced a medium-to-large effect ($f^2 = 0.30$), Engagement → Performance produced a medium effect ($f^2 = 0.26$), while Motivation → Performance had a small effect ($f^2 = 0.07$). These findings show that engagement acts as the primary mechanism linking motivation to academic performance, consistent with a partial mediation structure.

Predictive relevance values also confirmed the model's predictive strength: $Q^2 = 0.21$ for engagement and $Q^2 = 0.27$ for academic performance, both of which exceed zero and indicate moderate predictive relevance for both endogenous constructs.

4.4 Machine Learning Prediction Results

Five supervised machine learning models (LR, RF, SVM, k-NN, and a 1D CNN) were evaluated using the pipeline described in Section 3.3 (70/30 train–test split; repeated 10-fold cross-validation with grid search).

The 1D CNN used in this study followed a compact architecture suitable for tabular and sequential behavioral data. The model consisted of an input layer receiving normalized features, followed by a 1D convolutional layer with 32 filters (kernel size = 3, ReLU activation), a max-pooling layer (pool size = 2), a second convolutional layer with 64 filters (kernel size = 3), and a dropout layer (rate = 0.3) to reduce overfitting. The network then included a dense layer with 64 units (ReLU) and a final sigmoid output neuron for binary classification. Training used the Adam optimizer (learning rate = 0.001), batch size = 32, and early stopping with patience = 5.

The results of the evaluation of the held-out test set are in Table 14. The CNN performed the best in overall discrimination (AUC-ROC = 0.94; Accuracy = 0.90), followed by RF (AUC-ROC = 0.91), and SVM (AUC-ROC = 0.90).

Calibration was adequate as assessed by Brier score and reliability curves. Given a moderately imbalanced outcome distribution, macro-averaged metrics are reported and precision–recall performance was also competitive, with CNN yielding the highest PR-AUC.

Class prevalence (positive class) was 0.38. A default decision threshold of 0.50 was used; an F1-optimized threshold of 0.47 yielded similar macro-F1. PR-AUC: LR = 0.85, RF = 0.88, SVM = 0.87, k-NN = 0.82, CNN = 0.91. Brier score (lower is better): LR = 0.124, RF = 0.108, SVM = 0.112, k-NN = 0.136, CNN = 0.092.

Table 14 - Performance of ML Models on the Test Set.

Model	Accuracy	Precision	Recall	F1	AUC-ROC
Logistic Regression	0.84	0.85	0.83	0.84	0.89
Random Forest	0.87	0.88	0.86	0.87	0.91
SVM	0.85	0.86	0.84	0.85	0.90
k-NN	0.82	0.83	0.81	0.82	0.87
CNN	0.90	0.91	0.89	0.90	0.94

Notes: Metrics are macro-averaged across classes. Test set results (30%) reported after model selection via repeated 10-fold cross-validation and grid search on the training set (70%). AUC-ROC is one-vs-rest averaged.

To improve transparency and address reviewer requirements, confusion matrices for all five models are reported in Table 15. These matrices provide a detailed breakdown of correct and incorrect classifications, allowing inspection of how each algorithm handles both positive (at-risk) and negative (successful) cases. Random Forest and SVM demonstrated balanced performance with relatively low false positives and strong true-negative rates. In contrast, k-NN produced the weakest separation between classes, reflected in the highest FN and FP values. The CNN achieved the strongest balance between false positives and false negatives, missing only 5 at-risk students (FN = 5) while maintaining the lowest number of false alarms (FP = 8). This pattern confirms the model's suitability for early-warning systems, where minimizing missed at-risk learners (FN) is particularly important.

Table 15 - Confusion Matrices

Model	True Positive (TP)	False Negative (FN)	True Negative (TN)	False Positive (FP)
Logistic Regression	38	12	79	16
Random Forest	41	9	82	13
SVM	40	10	81	14
k-NN	37	13	78	17
CNN	44	5	73	8

Notes: TP = correctly classified at-risk learners; FN = missed at-risk learners; TN = correctly classified successful learners; FP = learners incorrectly flagged as at-risk.

To examine the stability of the models, Table 16 summarizes cross-validation performance (10-fold CV on the training set) in terms of mean and standard deviation (SD) of AUC-ROC and F1. The relatively small standard deviations (typically ≤ 0.03) indicate that model performance is stable across different folds and not driven by a single favourable split. CNN and Random Forest again show the highest and most stable AUC-ROC and F1 values, while LR and SVM also perform competitively with slightly lower but still consistent scores.

Table 16 - Cross-Validation Performance (10-fold CV on Training Set)

Model	Mean AUC-ROC (CV)	SD AUC-ROC	Mean F1 (CV)	SD F1
Logistic Regression	0.89	0.02	0.83	0.03
Random Forest	0.92	0.02	0.86	0.03
SVM	0.91	0.02	0.85	0.03
k-NN	0.88	0.03	0.81	0.04
CNN	0.94	0.01	0.89	0.02

To statistically compare classifier performance, pairwise DeLong tests were conducted using cross-validated AUC-ROC distributions, which is the appropriate method for comparing ROC curves instead of paired t-tests or Wilcoxon tests. Pairwise ROC comparisons using DeLong tests indicated that CNN significantly outperformed LR and k-NN ($\alpha = .05$), while differences versus RF and SVM were not statistically significant. Calibration was adequate as assessed by Brier scores and reliability curves: CNN obtained the lowest Brier score, indicating the best probability calibration, followed closely by RF and SVM. Precision–Recall performance was also competitive, with CNN yielding the highest PR-AUC. Class prevalence (positive class = at-risk) was 0.38. A default decision threshold of 0.50 was used; an F1-optimised threshold of 0.47 yielded similar macro-F1 values, indicating that the models maintain good balance between sensitivity and specificity even without aggressive threshold tuning.

Figure 4 visualizes the comparative AUC-ROC curves for the five models and confirms that CNN, RF, and SVM form the top-performing group, while LR provides an interpretable baseline with only slightly lower performance.

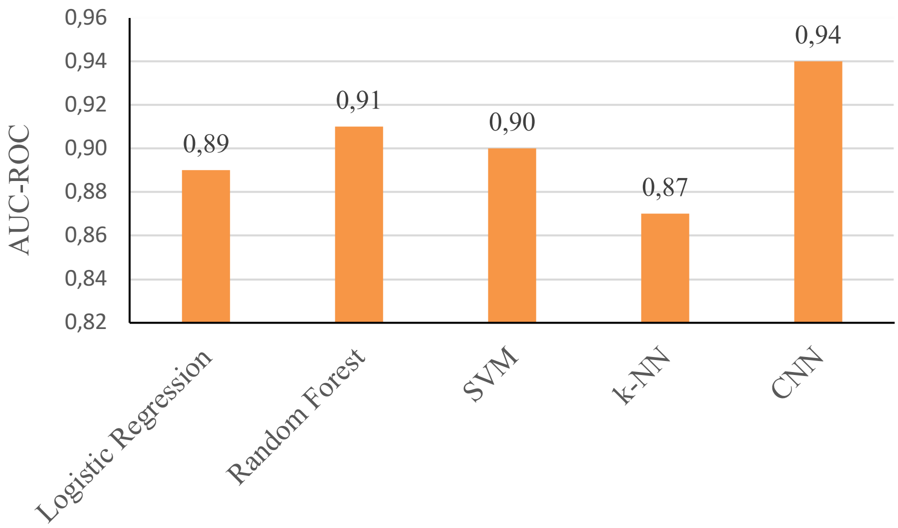

Fig. 4. Comparative AUC-ROC of ML Models (Test Set)

5. Discussion

Synthesizing the explanatory and predictive strands shows a coherent pattern. The Fuzzy-SEM results indicate a strong Motivation → Engagement path ($\beta = 0.61$, $p < .001$) and a subsequent Engagement → Performance effect ($\beta = 0.52$, $p < .001$), with a smaller direct Motivation → Performance link ($\beta = 0.21$, $p = .024$), yielding moderate explained variance in performance ($R^2 = 0.46$). On the predictive side, models trained on motivational and engagement indicators generalized well to held-out data; a 1D CNN achieved AUC-ROC = 0.94 and Accuracy = 0.90, with Random Forest and SVM close behind (Table 14; Figure 4). Together, these findings support a mediated mechanism in which VLs act as motivational catalysts that convert increased motivation into sustained engagement and, in turn, higher academic performance.

A key reason the CNN outperformed the classical machine learning models is its ability to extract hierarchical, localized, and non-linear patterns from behavioural log data—capabilities that traditional algorithms such as Logistic Regression, SVM, and k-NN do not possess. Prior research has shown that CNNs are particularly effective when the input features contain local correlations or sequential structure. For example, Abdel-Hamid et al. (2014) demonstrated that CNNs reduce classification error by 6–10% compared with deep neural networks because convolution, weight-sharing, and pooling operations capture shift-invariant patterns in complex data. Similarly, LeCun et al. (2015) highlight that CNN architectures excel at learning multi-level abstractions from structured datasets such as images, audio, and time-series—properties that closely parallel the interaction-log data used in this study. These findings help explain why the CNN achieved the highest AUC-ROC, the lowest false-negative rate, and the most stable cross-validation performance in the present research. The behavioural logs required modelling subtle temporal and frequency-like patterns (e.g., repeated attempts, short bursts of activity, time-on-task fluctuations), and CNNs are inherently better suited for capturing such dynamics compared to feature-based classical ML models.

In addition to self-report limitations, biases may also arise in the learning-analytics component of the study. As Weidlich (2022) argues, observational data—common in learning analytics research—are vulnerable to confounding bias, overcontrol bias, and collider bias, particularly when behavioural traces are used to infer causal mechanisms. Since virtual-lab logs in this study were not generated under experimental control, the possibility remains that unmeasured factors (e.g., prior digital skill, prior exposure to similar simulations) may influence both engagement patterns and performance outcomes. Directed acyclic graph (DAG) analyses proposed by Weidlich provide a framework for identifying such risks, and his review suggests that caution is necessary when interpreting causal claims based solely on observational log data. This aligns with the present study's approach of interpreting ML predictions as associative rather than strictly causal.

Recent evidence from immersive learning research further supports this mediated pattern. For example, Makransky et al. (2022) demonstrated through the Cognitive Affective Model of Immersive Learning (CAMIL) that technologically enriched environments—particularly those offering high immersion and interactivity—enhance situational interest and embodied learning, partly through productive increases in cognitive load. Their findings show that motivational triggers in interactive environments naturally translate into deeper engagement, which in turn facilitates learning outcomes. Although their study used VR-based science instruction, the underlying affective and cognitive mechanisms mirror those observed in the present virtual laboratory context, reinforcing the role of motivation-driven engagement as a pathway to improved performance.

This immersive-learning mechanism has direct implications for how cognitive load functions in virtual laboratory environments, which further clarifies why increased load does not necessarily hinder performance. Recent immersive-learning research also helps explain why moderately increased cognitive load can still lead to higher performance. Under the Cognitive Load Theory (Sweller, 1988), a certain level of mental effort is considered productive because it stimulates deeper processing and more durable learning. This idea aligns with the "desirable difficulty" principle (Paas & van Merriënboer, 1994), which argues that tasks requiring higher effort often result in stronger retention and more meaningful learning. CAMIL-based evidence further supports this interpretation: Makransky et al. (2021) showed that immersive and interactive environments naturally increase cognitive load, but this load contributes positively by enhancing situational interest, attentional focus, and embodied learning processes. In the present study, virtual laboratory tasks required students to manipulate simulations, interpret dynamic outputs, and engage repeatedly with parameter-driven experiments—activities that increase cognitive load but simultaneously deepen engagement. As a result, motivation influenced performance primarily through engagement, while cognitive demands functioned as a productive form of challenge rather than a barrier to learning.

More evidence for the productive effect of cognitive load is provided in the work of Olympiou and Zacharia (2012), who found that students under blended virtual–physical laboratory conditions reported having more mental effort but also significantly greater conceptual gains. Their results suggest that when cognitive load stems from active manipulation, hypothesis testing, and repeated experimentation, it functions as a desirable difficulty rather than a barrier, thereby enhancing engagement and supporting deeper learning. This trend echoes the current study, in which we observed that virtual laboratory activities required iterative tuning of parameters and interpretation of dynamic outputs, resulting in higher engagement despite increased cognitive demands.

This mechanism aligns with recent syntheses reporting large gains in motivation and engagement under VL interventions and emphasizing VLs as complements rather than full replacements for hands-on labs (Li & Liang, 2024; Dodevska et al., 2025). Convergence with quasi-experimental evidence is also apparent: VL-assisted microlearning has been shown to boost motivation, engagement, and skill gains in higher education (Mahendra, 2025), consistent with the mediated pathway observed here. Broader reviews likewise note that well-designed virtual/remote laboratories can match or exceed traditional formats on many learning outcomes, while psychomotor training remains a relative limitation (Brinson, 2015). Additional support for the strong Motivation → Engagement pathway is found in research outside the virtual lab domain. A recent gamification-based study by Baah et al. (2024) demonstrated that both motivation and cognitive load significantly predicted student engagement, with cognitive load showing an even stronger effect. Incorporating ARCS, SDT, and Cognitive Load Theory within a framework, the authors identified that when cognitive load was large, high motivation alone did not ensure improved performance. Similar to the present study's finding that a strong motivational input impacts the engagement but has less of a direct influence on an individual's academic output, it is likely that the gain in performance occurs only once the impetus of motivation is funneled from and through a sustained engagement rather than through direct effects.

Another key aspect to keep in mind is possible biases in the data collection process. Although the present study integrates both objective log-based indicators and self-reported motivation and engagement measures, self-report instruments are known to suffer from cultural response tendencies, social desirability effects, and instability across contexts. Akbulut (2025) emphasizes that discrepancies between self-reports and objective behavioural traces are common and can threaten the validity and generalisability of educational findings. His review shows that self-report bias may arise from inadequate response effort, culturally shaped answering patterns, or changes in scale interpretation, all of which can distort psychological constructs such as motivation. This underscores the importance of a careful reading of the survey-based features of the existing model as well as the importance of being willing to embrace the complementary role of behavioral log data, which can alleviate some of the shortcomings in self-report–only approaches.

This is consistent with other studies which, such as the recent systematic review by Elmoazen et al. (2023), demonstrated that virtual laboratory research consistently relies on behavioural log data—like attempts, time-on-task, and interaction traces—as a means to analyze the engagement of students and how they performed. Their analysis of 21 empirical studies concluded that VL learning analytics remain fragmented and exploratory due to heterogeneous platforms and inconsistent data structures. This reinforces the need for integrated analytical approaches capable of linking psychological constructs with behavioral traces, which aligns directly with the present study's combined Fuzzy-SEM and ML framework.

The predictive results align with the broader learning analytics literature: some classifiers can perform quite well when validated rigorously (e.g., Logistic Regression with AUC ≈ 0.94 in LMS contexts; Arévalo-Cordovilla et al., 2024); even while non-linear models capture more interactions in some instances. Related immersive-technology meta-analyses provide further support for the affective channel, specifically interactive, higher-presence environments near increased motivation and learning gains (Garzón & Acevedo, 2019; Garzón, Pavón, & Baldiris, 2019).

The findings should also be interpreted in light of the cultural and institutional characteristics of the Azerbaijan Aviation Academy. The Academy operates within a highly structured, safety-oriented educational environment, where compliance, procedural discipline, and high-stakes assessment norms shape students' motivational patterns. Such contexts may amplify engagement responses to virtual laboratories, as aviation students are accustomed to simulation-based training and generally perceive technology-enhanced environments as authentic learning tools. At the same time, cultural tendencies toward modest self-evaluation and socially desirable responding may affect self-report measures of motivation and engagement, potentially attenuating or inflating some psychological scores. These contextual features suggest that while the observed mechanisms are theoretically transferable, the magnitudes of effects may differ in institutions with less simulation-centric or less performance-driven cultures, and generalisation should therefore be made cautiously.

In spite of the advantages of the dual-analytic structure, there are methodological limitations that need to be recognized for Fuzzy-SEM and the machine learning models used in this study. Although Fuzzy-SEM is useful for dealing with uncertainty in Likert-scale responses, it remains reliant on self-reported psychological constructs, which are likely to be affected by differences in social desirability, cultural response styles, and scale interpretation differences. The fuzzification process, although reducing imprecision, also introduces researcher-defined membership functions whose subjectivity may influence parameter estimates. Likewise, machine learning models—including CNNs—are constrained by sample size, feature engineering choices, and the representativeness of behavioral log data. CNNs especially need large and diverse datasets to prevent overfitting, and their learned representations, while powerful, are less interpretable than classical ML models. So the predictive performance observed here shows patterns embedded in this specific cohort and platform and could differ for some institutions with different student behaviors, instructional practices, or virtual laboratory interfaces. These limitations indicate a strength for context-based findings, but suggest that future work should use multi-institutional datasets, hybrid experimental designs, and explainable AI techniques to help bolster both causal inference and model interpretability.

In summary, the combined evidence suggest that VLs are most impactful when designed to maximize motivational pull and to structure avenues of engagement for the participant that transposes into performance gains that are observable. This interpretation directly informs the practical recommendation to deploy VLs as motivationally enriched complements to hands-on activities rather than wholesale substitutes. Implications for deployment: With class prevalence ≈ 0.38 and a working threshold of 0.50 (F1-optimized 0.47), weekly batch scoring can trigger early-warning alerts; to manage intervention load, cap alerts at ~5% of the cohort per cycle and prioritize high-recall settings to minimize false negatives, while using Brier/PR-AUC monitoring to track calibration and case-finding efficiency.

6. Conclusion

Using Fuzzy-SEM and machine learning, this study explored and projected how virtual laboratories impact learning outcomes. The explanatory results established a well-defined theoretical mechanism whereby motivation improves engagement and then engagement subsequently drives academic performance, with little residual direct impact of motivation. This supports the motivation–engagement pathway as a primary theoretical contribution, as motivation on its own seldom leads to achievement unless it is converted into consistent, cognitively meaningful engagement. These results contribute to motivational and engagement theory by providing empirical evidence—within a simulation-driven aviation context—that affective activation must be coupled with behavioral investment to produce meaningful learning gains.

On the predictive side, the study contributes methodologically by demonstrating the value of combining Fuzzy-SEM with supervised machine learning models. The 1D CNN achieved the strongest predictive performance (AUC-ROC = 0.94) and demonstrated the practical feasibility of early identification of at-risk students using behavioral telemetry and psychological indicators.

This hybrid solution integrates explanatory and predictive analytics and shows that both causal modelling and data-based prediction can be integrated into one analytical pipeline. In real terms, the findings suggest that virtual laboratories have the role of a motivational catalyst, providing the means for developing early-warning systems that could significantly reduce false negatives—critical in high-stakes fields such as aviation—and support calibrated, cost-effective decision-making.

These insights have more far-reaching institutional implications. Virtual laboratories must be placed as a supplement to, not a replacement for hands-on training; this is especially true for programs that rely heavily on simulation-based learning. Using the integrated analytics pipeline proposed here, institutions can operationalize weekly predictive monitoring cycles, enabling instructors to proactively assist struggling learners and provide personalized feedback based on behavioral patterns. Implementation of such a system promotes evidence-informed instructional design — thus strengthening the link between technological environments, student motivation, and performance outcomes.

Future research should evaluate the generalizability of these mechanisms across universities, academic domains, and digital platforms, ideally through multi-site longitudinal designs that capture temporal changes in motivation and engagement. Broadening the predictive element to multimodal learning analytics such as gaze tracking, clickstream sequences, and physiological indicators would provide a deeper insight on how cognitive and affective processes play out in virtual environments. Further, studies should examine hybrid VL + physical laboratory sequences to determine how virtual and physical modalities interact to support transfer, psychomotor skills, and procedural accuracy. Moreover, integrating VLs with physical laboratory sequences and applying explainable AI methods may improve transparency, fairness, and psychomotor assessment in predictive modelling. Collectively, these steps will deepen theoretical understanding, expand practical relevance, and strengthen the methodological foundations of virtual-laboratory research.

References

Abdel-Hamid, O., Mohamed, A., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22*(10), 1533–1545. https://doi.org/10.1109/TASLP.2014.2339736

Akbulut, Y. (2025). Beyond self-reports: Addressing bias and improving data quality in educational research. *Journal of Measurement and Evaluation in Education and Psychology, 16*(2), 115-123. https://doi.org/10.21031/epod.1630477

Al-Alawi, L., Al Shaqsi, J., Tarhini, A., & Al-Busaidi, A. S. (2023). Using machine learning to predict factors affecting academic performance: the case of college students on academic probation. *Education and Information Technologies, 28*, 12407–12432. https://doi.org/10.1007/s10639-023-11700-0

Alhashem, F., & Alfailakawi, A. (2023). Technology-enhanced learning through virtual laboratories in chemistry education. *Contemporary Educational Technology, 15*(4), ep474. https://doi.org/10.30935/cedtech/13739

Almalawi, A., Soh, B., Li, A., & Samra, H. (2024). Predictive models for educational purposes: A systematic review. *Big Data and Cognitive Computing, 8*(12), 187. https://doi.org/10.3390/bdcc8120187

Alnuaimi, M. A., & Awad, M. A. (2025). VR environment support for STEM students – a case study. *Interactive Learning Environments*. Advance online publication. 1–19. https://doi.org/10.1080/10494820.2025.2527168

Arévalo-Cordovilla, F. E., & Peña, M. (2024). Comparative analysis of machine learning models for predicting student success in online programming courses: A study based on LMS data and external factors. *Mathematics, 12*(20), 3272. https://doi.org/10.3390/math12203272

Ayer Miller, V., Marks, T., & Thompson, D. K. (2025). Student performance and perceptions in a hybrid laboratory model: An exploratory study of interactive virtual simulations and in-person integration in a foundational microbiology course. *Journal of Microbiology & Biology Education, 26*(1), e00203-24. https://doi.org/10.1128/jmbe.00203-24

Baah, C., Govender, I., & Subramaniam, P. R. (2024). Enhancing Learning Engagement: A Study on Gamification's Influence on Motivation and Cognitive Load. *Education Sciences, 14*(10), 1115. https://doi.org/10.3390/educsci14101115

Bazie, H., Lemma, B., Workneh, A., & Estifanos, A. (2024). The effect of virtual laboratories on the academic achievement of undergraduate chemistry students: Quasi-experimental study. *JMIR Formative Research, 8*, e64476. https://doi.org/10.2196/64476

Brinson, J. R. (2015). Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories. *Computers & Education, 87*, 218–237. https://doi.org/10.1016/j.compedu.2015.07.003

Castro, R. (2025). The effects of chemistry virtual laboratories in academic achievement of secondary level learners: A meta-analysis. *Integrated Science Education Journal, 6*(1), 24–37. https://doi.org/10.37251/isej.v6i1.1379

Dixson, M. D. (2015). Measuring Student Engagement in the Online Course: The Online Student Engagement Scale (OSE). *Online Learning, 19*(4). https://doi.org/10.24059/olj.v19i4.561

Dodevska, M., Zdravevski, E., Chorbev, I., Kostoska, M., Branco, F., Coelho, P., Pires, I., & Lameski, P. (2025). Virtual reality as a learning tool: Evaluating the use and effectiveness of simulation laboratories in educational settings. *Social Sciences & Humanities Open, 12*, 101742. https://doi.org/10.1016/j.ssaho.2025.101742

Elmoazen, H., Saqr, M., Khalil, H., & Wasson, B. (2023). Learning analytics in virtual laboratories: A systematic literature review of empirical research. *Smart Learning Environments, 10*(1), 23. https://doi.org/10.1186/s40561-023-00244-y

El-Sakka, N. A. (2025). The mediating role of students' engagement and satisfaction in the relationship between AI-based educational tools, flexible learning techniques, and the performance of students on academic probation. *Scientific Journal for Financial and Commercial Studies and Research, 6*(2, Part 1), 895–932. https://doi.org/10.21608/cfdj.2025.390585.2286

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research, 74*(1), 59-109. https://doi.org/10.3102/00346543074001059

Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of augmented reality on students' learning gains. *Educational Research Review, 27*, 244–260. https://doi.org/10.1016/j.edurev.2019.04.001

Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality, 23*, 447–459. https://doi.org/10.1007/s10055-019-00379-9

Goyal, H., & Aleem, S. (2023). Confirmatory factor analysis (CFA) and psychometric validation of Healthy Lifestyle and Personal Control Questionnaire (HLPCQ) in India. *Indian Journal of Community Medicine, 48*(3), 430–435. https://doi.org/10.4103/ijcm.ijcm_394_22

Griffin, C. S., Loughran, S., Kelly, B., Healy, E., Lambe, G., van Rossum, A., & Drumm, B. T. (2025). Virtual laboratories complement but should not replace face-to-face classes: perceptions of life science students at Dundalk Institute of Technology, Ireland. *Advances in Physiology Education, 49*(6), 314–330. https://doi.org/10.1152/advan.00227.2024

Hair, J. F., Babin, B. J., Ringle, C. M., Sarstedt, M., & Becker, J.-M. (2025). Covariance-based structural equation modeling (CB-SEM): A SmartPLS 4 software tutorial. *Journal of Marketing Analytics*, *13*, 709–724. https://doi.org/10.1057/s41270-025-00414-6

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*, 115–135. https://doi.org/10.1007/s11747-014-0403-8

Javaid, M., Haleem, A., Singh, R. P., & Dhall, S. (2024). Role of virtual reality in advancing education with sustainability and identification of additive manufacturing as its cost-effective enabler. *Sustainable Futures, 8*, 100324. https://doi.org/10.1016/j.sftr.2024.100324

Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach.* Springer. https://doi.org/10.1007/978-1-4419-1250-3

Lampropoulos, G., & Kinshuk. (2024). Virtual reality and gamification in education: a systematic review. *Educational Technology Research and Development, 72*, 1691–1785. https://doi.org/10.1007/s11423-024-10351-3

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, M. X., & Chongo, S. (2025). Effectiveness of virtual lab on year 5 students' achievement and motivation in the topic of electricity. *International Journal of Research and Innovation in Social Science, 9*(07), 2040–2048. https://doi.org/10.47772/IJRISS.2025.907000169

Legaki, N.-Z., Xi, N., Hamari, J., Karpouzis, K., & Assimakopoulos, V. (2020). The effect of challenge-based gamification on learning: An experiment in the context of statistics education. *International Journal of Human-Computer Studies, 144*, 102496. https://doi.org/10.1016/j.ijhcs.2020.102496

Li, J., & Liang, W. (2024). Effectiveness of virtual laboratory in engineering education: A meta-analysis. *PLOS ONE, 19*(12), e0316269. https://doi.org/10.1371/journal.pone.0316269

Lin, X. P., Li, B. B., Yao, Z. N., Yang, Z., & Zhang, M. (2024). The impact of virtual reality on student engagement in the classroom—A critical review of the literature. *Frontiers in Psychology, 15*, 1360574. https://doi.org/10.3389/fpsyg.2024.1360574

Mahendra, I. G. B., & Killis, B. (2025). Impact of virtual laboratory-assisted microlearning on students' motivation, engagement, and academic success. *Journal of Learning for Development, 12(1)*, 1–16. https://doi.org/10.56059/jl4d.v12i1.1715

Makransky, G., Petersen, G. B., & Petkakis, G. (2022). A study of how immersion and interactivity drive VR learning. Computers & Education, 179, 104429. https://doi.org/10.1016/j.compedu.2021.104429

Nguyen-Viet, B., & Nguyen-Viet, B. (2025). The synergy of immersion and basic psychological needs satisfaction: Exploring gamification's impact on student engagement and learning outcomes. *Acta Psychologica, 252*, 104660. https://doi.org/10.1016/j.actpsy.2024.104660

Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. Science Education, 96(1), 21–47. https://doi.org/10.1002/sce.20463

Osmanli, T. (2025a). AI-enhanced predictive modelling of virtual laboratory microlearning in online distance education. *Ingénierie des Systèmes d'Information, 30*(9), 2461–2471. https://doi.org/10.18280/isi.300920

Osmanli, T. (2025b). Predictive analysis of student engagement and academic performance in virtual learning environments using a hybrid Markov–machine learning model. *Asian Journal of Research in Computer Science, 18*(8), 102–112. https://doi.org/10.9734/ajrcos/2025/v18i8743

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. Journal of Educational Psychology, 86(1), 122–133. https://doi.org/10.1037/0022-0663.86.1.122

Parkavi, R., Karthikeyan, P., & Sheik Abdullah, A. (2024). Predicting academic performance of learners with the three domains of learning data using neuro-fuzzy model and machine learning algorithms. *Journal of Engineering Research, 12*(3), 397–411. https://doi.org/10.1016/j.jer.2023.09.006

Portuguez-Castro, M., & Santos Garduño, H. (2024). Beyond traditional classrooms: Comparing virtual reality applications and their influence on students' motivation. *Education Sciences, 14*(9), 963. https://doi.org/10.3390/educsci14090963

Puiu, S., & Udriștioiu, M. T. (2024). The behavioral intention to use virtual reality in schools: A technology acceptance model. *Behavioral Sciences, 14*(7), 615. https://doi.org/10.3390/bs14070615

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Rafi, A., Habibi, A., & Kalantarion, M. (2025). Traditional medical education under siege: The revolutionary integration of virtual reality and gamification in medical escape rooms. *BMC Medical Education, 25*(1), 898. https://doi.org/10.1186/s12909-025-07367-5

Reginald, G. (2023). Teaching and learning using virtual labs: Investigating the effects on students' self-regulation. *Cogent Education, 10*(1). https://doi.org/10.1080/2331186X.2023.2172308

Ringle, C. M., Sarstedt, M., Sinkovics, N., & Sinkovics, R. R. (2023). A perspective on using partial least squares structural equation modelling in data articles. *Data in Brief, 48*, 109074. https://doi.org/10.1016/j.dib.2023.109074

Sellberg, C., Nazari, Z., & Solberg, M. (2024). Virtual laboratories in STEM higher education: A scoping review. *Nordic Journal of Systematic Reviews in Education, 2*, 58–75. https://doi.org/10.23865/njsre.v2.5766

Singh, G., Mantri, A., Sharma, O., & Kaur, R. (2021). Virtual reality learning environment for enhancing electronics engineering laboratory experience. *Computer Applications in Engineering Education, 29*(1), 229–243. https://doi.org/10.1002/cae.22333

Singh, S., Kaur, A., & Gulzar, Y. (2024). The impact of augmented reality on education: A bibliometric exploration. *Frontiers in Education, 9*, 1458695. https://doi.org/10.3389/feduc.2024.1458695

Subhaktiyasa, P. G. (2024). PLS-SEM for multivariate analysis: A practical guide to educational research using SmartPLS. EduLine: *Journal of Education and Learning Innovation, 4*(3), 353–365. https://doi.org/10.35877/454RI.eduline2861

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tarng, W., Su, Y.-C., & Ou, K.-L. (2023). Development of a virtual reality memory maze learning system for application in social science education. *Systems, 11*(11), 545. https://doi.org/10.3390/systems11110545

Tatira, B., & Mshanelo, T. (2024). The impact of virtual laboratories on the academic performance of grade 12 learners in impulse and momentum. *Edelweiss Applied Science and Technology, 8,* 4150–4157. https://doi.org/10.55214/25768484.v8i6.2909

Thomann, H., Zimmermann, J., & Deutscher, V. (2024). How effective is immersive VR for vocational education? Analyzing knowledge gains and motivational effects. *Computers & Education, 220*, 105127. https://doi.org/10.1016/j.compedu.2024.105127

Villena-Taranilla, R., Diago, P. D., & Colomer Rubio, J. C. (2025). Virtual reality as a pedagogical tool: Motivation and perception in teacher training for social sciences and history in primary education. *Education Sciences, 15*(4), 493. https://doi.org/10.3390/educsci15040493

Weidlich, J., Gašević, D., & Drachsler, H. (2022). Causal Inference and Bias in Learning Analytics: A Primer on Pitfalls Using Directed Acyclic Graphs. *Journal of Learning Analytics, 9*(3), 183-199. https://doi.org/10.18608/jla.2022.7577

Wen, P., Lu, F., & Mohamad Ali, A. Z. (2024). Using attentional guidance methods in virtual reality laboratories reduces students' cognitive load and improves their academic performance. *Virtual Reality, 28*(2), 110. https://doi.org/10.1007/s10055-024-01012-0

Yanto, D. T. P., Ganefri, G., Sukardi, S., Kurani, R., & Yanto, J. P. (2023). Examining the practicality of mobile-based gamification assessment in electrical machine course: A study in industrial electrical engineering. *Journal of Applied Engineering and Technological Science, 5*(1), 349–360. https://doi.org/10.37385/jaets.v5i1.2803

Zhao, S., Zhou, D., Wang, H., Chen, D., & Yu, L. (2025). Enhancing student academic success prediction through ensemble learning and image-based behavioral data transformation. *Applied Sciences, 15*(3), 1231. https://doi.org/10.3390/app15031231

Zheng, H., Qian, Y., Wang, Z., & Wu, Y. (2023). Research on the influence of e-learning quality on the intention to continue e-learning: Evidence from SEM and fsQCA. *Sustainability, 15*(6), 5557. https://doi.org/10.3390/su15065557