British Journal of Education, 14 (6),58-77, 2026
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

Integrating Partial Least Squares Structural Equation Modelling and Artificial Intelligence in Educational Research: A Simulation-Based Methodological Framework for Resource-Constrained Contexts

Caroline Ochuko Alordiah
Faculty of Education, University of Delta, Agbor, Nigeria
ORCID ID: https://orcid.org/0000-0002-2917-9302

John Oji
Faculty of Education, University of Delta, Agbor, Nigeria

doi: https://doi.org/10.37745/bje.2013/vol14n65877 Published July 11, 2026

Citation: Alordiah C.O. and Oji J. (2026) Integrating Partial Least Squares Structural Equation Modelling and Artificial Intelligence in Educational Research: A Simulation-Based Methodological Framework for Resource-Constrained Contexts, *British Journal of Education*, 14 (6),58-77

Abstract: *The rise in the need to make decisions in the educational system based on data has led to the application of sophisticated data analysis methods like Partial Least Squares Structural Equation Modelling (PLS-SEM) and Artificial Intelligence (AI). Nevertheless, their joint use is not yet widespread, especially in resource-limited scenarios. This paper suggests and illustrates a unified methodological framework of applying PLS-SEM and AI via a simulation-based method. The explanatory modelling and machine learning (Random Forest) of synthetic data to represent key educational constructs were performed using PLS-SEM and machine learning, respectively. The results of the PLS-SEM showed that there were strong correlations between constructs (satisfaction, student engagement, and institutional support) that described 54% of the variance in academic performance. The AI model had a better predictive accuracy (R^2 = 0.71), as compared to the PLS-SEM model. The integration had complementary and consistent results. The research shows that the combination of PLS-SEM and AI increases the explanatory and predictive power. The framework is specifically applicable to resource-limited settings and provides viable advice to education researchers.*

Keywords: artificial intelligence, education research, PLS-SEM, predictive analytics, resource-constrained contexts, simulation

INTRODUCTION

The growing movement toward using data-driven decision-making in the education sector has served to drive the uptake of sophisticated analytical tools that are able to encompass a variety of relationships and produce insights that can be put into action. Out of these methods, Partial Least Squares Structural Equation Modelling (PLS-SEM) and Artificial Intelligence (AI) have become widely used methods of analysing educational data. PLS-SEM is known to provide the capability of modelling latent constructs and testing theoretical relationships, whereas AI methods are good at predictive modelling and pattern recognition (Hair et al., 2019; Richter & Tudoran, 2024; Zawacki-Richter et al., 2019). Although these methods have their own advantages, they are mostly used in isolation and their ability to give detailed information with both explanatory and predictive capabilities is curtailed (Ayad, 2025; Oji and Alordiah, 2024; Magasi, 2025).

In the field of education research, it is important to analyse the relationships between constructs like student engagement, student satisfaction, institutional support and academic performance to enhance the learning process and policy making (Enokela et al., 2026). The flexibility of PLS-SEM and limited distributional assumptions as well as its appropriateness in small sample sizes have also made it highly applicable in modeling these relationships (Guenther et al., 2023; Hair et al., 2019; Haji-Othman et al., 2024). Nevertheless, it has a relatively narrow focus on explanation and causal inference, and thus it might not be taking full advantage of the potential of educational data prediction (Demir & Uşak, 2025; Subhaktiyasa, 2024). On the other hand, AI methods, such as machine learning models like random forests and neural networks, are capable of powerful prediction but are not always interpretable, which is why it can be hard to identify how such predictions are obtained (Haddouchi & Berrado, 2024; Holmes et al., 2019; Scorzato, 2024).

It is the rise of interest in combining theory-driven and data-driven methods that has been fostered by the growing need to combine the two components of interpretability and predictive accuracy. New methodological developments underline the significance of using explanatory modelling and predictive analytics in unison to increase the strength and practical relevance of research results (Hair, J., & Sarstedt, 2021; Richter & Tudoran, 2024; Shmueli et al., 2019; Wan & Wan, 2023). In this regard, incorporating both PLS-SEM and AI is a promising avenue to overcome the gap between the two paradigms. This type of integration helps researchers test theoretical models and, at the same time, apply data-driven methods to enhance the accuracy of predictions.

The necessity is especially high in resource-limited settings, where analysts are confronted with insufficient data access, insufficient computational resources, and limited access to state-of-the-art analytical software. Such limitations tend to make it difficult to apply advanced analytical methods and can lead to differences in research capacities in different regions. As such, methodological frameworks that are, not only, robust, but flexible to such contexts are needed.

The study fills this gap by suggesting and illustrating a unified methodological approach to education research using PLS-SEM and AI. In contrast to strictly conceptual research, the current study uses a simulation approach to show the practical application of the framework, under controlled, but realistic conditions. Simulation allows one to reproduce the common data structures and relationships found in education research, thus offering a valid foundation to demonstrate the methodology in case there is not enough or no real-world data available (Sarstedt et al., 2020; Sturman, 2023).

Although there is increasing interest in both PLS-SEM and artificial intelligence, the current body of research mostly uses these methods separately without much methodological advice on how to integrate them, especially in the case of education research (Kusumah et al., 2024). In addition, the previous research lacked the sufficient coverage of the way in which such integration can be operationalized within the context of resource-constrained conditions. This paper contributes to the literature by suggesting a framework of PLS-SEM and AI simulation and step-by-step integration, which combines theoretical substantiation with predictive modelling. In contrast to previous research, this article shows that the latent variable scores derived with the PLS-SEM could be successfully used as an input in machine learning models, thus, increasing the interpretability and predictive capabilities in a single analytical setting.
It is based on this that the study aims to provide answers to the following research question:
What are the ways of utilizing PLS-SEM and AI to complement each other to better clarify and predict in the field of educational research in resource-constrained settings?

In order to answer this question, the research aims at achieving the following:
(1) to create a conceptual and methodological framework to unify PLS-SEM and AI;
(2) to show how the framework was implemented with simulated data;
(3) to compare the explanatory and predictive performance of the integrated approach; and
(4) to offer useful advice to researchers in resource-constrained settings.

LITERATURE REVIEW AND METHODOLOGICAL FRAMEWORK

PLS-SEM Research in Education.

Partial Least Squares Structural Equation Modelling (PLS-SEM) has become an widely accepted analytical method used to model the complex relationships among latent constructs in education research. PLS-SEM, in comparison with covariance-based SEM, is also a variance-based method and can be effectively used in exploratory studies, theory-building, and research with small sample sizes or non-normative data distributions (Hair et al., 2019; Hair et al., 2017). These features make it particularly applicable in the educational settings, where constraints in data and the changing theoretical frameworks are prevalent.

PLS-SEM has been widely used in the study of constructs like student engagement, attitude, behaviour, satisfaction, learning outcomes, and institutional effectiveness in the field of education research (Gutu et al., 2024; Iqbal et al., 2026; Singh et al., 2024). As an example, e-learning adoption research has often used PLS-SEM to examine behavioural intention and usage behaviour, especially in developing environments where technological infrastructure can be an uneven distribution (Al-Emran et al., 2018). Likewise, studies of student satisfaction and teaching efficacy tend to use PLS-SEM to prove the measurement models and evaluate the causal relationship between pedagogical variables (Ringle et al., 2020).

One of the main strengths of PLS-SEM is that it allows testing hypothesised relationships by simultaneously testing measurement and structural models, which assure reliability and validity of constructs. Also, hierarchical constructs and complex path relationships are typically modelled in the approach and are typical of education research (Sarstedt et al., 2020). Nevertheless, in spite of these benefits, PLS-SEM is more explanatory than explanatory. This makes it less effective in maximizing predictive accuracy, especially in the context where the forecast of outcomes (e.g., academic performance, student retention) is vital (Shmueli et al., 2019).

AI in Education Research.

Artificial Intelligence (AI) has greatly changed the sphere of education research by providing predictive analytics, adaptive learning environments, and data-driven decisions. Decision trees, random forests, support vector machines, and neural networks are examples of machine learning algorithms which are commonly used to analyse educational data and predict results (student performance, probability of dropping out, etc.). (Mohamed, 2025; Riyanto et al., 2024; Zawacki-Richter et al., 2019).

Learning analytics and educational data mining AI-based methods have shown good predictive power. As an example, machine learning can be used to detect at-risk students by analyzing behavioural and performance data, therefore, allowing early interventions (Nimy et al., 2023). Additionally, AI has enabled the creation of intelligent tutoring software and customized learning spaces, which respond to the needs of individual learners by adjusting educational content.

In spite of these advantages, AI techniques have a number of weaknesses. Lack of interpretability, commonly known as the black-box problem, is one of the greatest challenges. Numerous machine learning models generate correct predictions but little understanding of the determination of the variables, which is a barrier to their adoption in the teaching contexts where clarity is a prerequisite (Holmes et al., 2019). Also, AI models usually need large and quality datasets and in-depth computer computation power that might not be easily accessible in resource-constrained conditions.

PLS-SEM + AI: The Future of Explainable and Predictive Analytics.

The shortcomings of PLS-SEM and AI have given rise to a growing enthusiasm to combine these methods to utilize the synergistic advantages. PLS-SEM is more interpretable and theoretically based, whereas AI is more accurate in predictions and can model complex non-linear relationships. By integrating such approaches, researchers can strike a balance between prediction and explanation and, therefore, enhance the validity and usability of the research results in general.That integration is consistent with the new paradigm of explainable and predictive analytics, which insists on the significance of comprehending relations, as well as predicting the results correctly (Shmueli et al., 2019). A hybrid approach is especially useful in the context of education research, as it enables a researcher to both prove the theoretical models and produce any actionable predictions that can be used to make decisions.

One type of integration strategy is to use latent variable scores based on PLS-SEM as inputs in machine learning models. The methodology will result in a predictive model that is based on theoretically verified constructs, which will increase interpretability but predictive performance. Nonetheless, the application of PLS-SEM and AI is under-researched in the literature, especially regarding education research.

Although the potential of integrating explanatory and predictive approaches has been recognized in past studies, they tend to be conceptual with no practical implementation strategies. Specifically, the empirical or simulation-based evidence of how the outputs of PLS-SEM can be operationalized in AI models is limited. Besides, current research seldom addresses the limitations that researchers in developing contexts experience, including a lack of data and computing power. This shortcoming points to the necessity of a systematic and flexible framework that does not only combine both approaches but also gives definite indicators that should be followed in a resource limited setting.

Research Gap

Despite the widespread use of PLS-SEM and AI in the field of education research, the current research trends are inclined to employ the approaches separately which lacks systematic structures in which researchers can be guided on the way to integrate them. Moreover, there has not been much focus on the use of such integrated methodologies in resource-constrained settings where issues in data availability, computational power, and technical skills are common.

Moreover, most of the literature lacks practical implementation guidelines that offer step-by-step procedure of how to integrate PLS-SEM and AI. This gap restricts access to sophisticated methods of analysis to researchers, especially those working in developing settings.

Methodological Framework

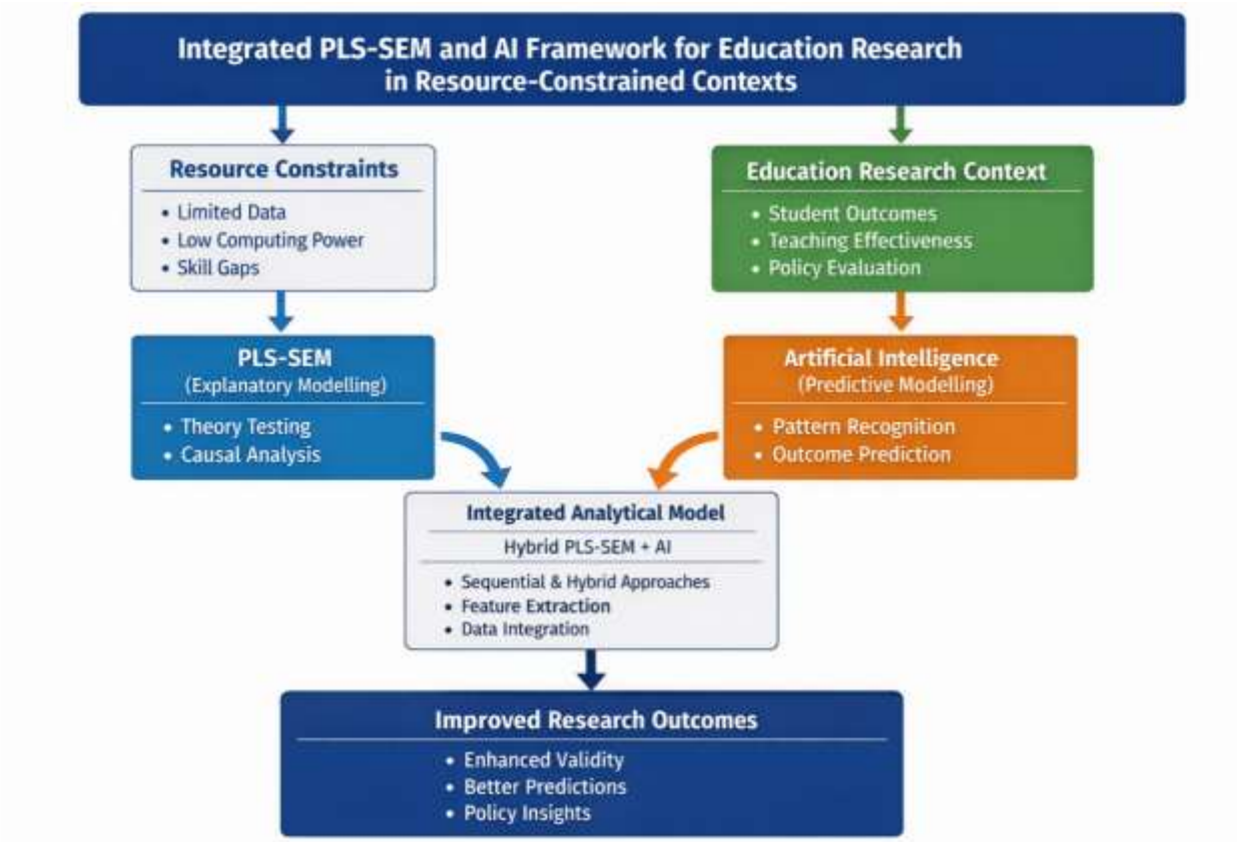

Figure 1. Conceptual framework integrating PLS-SEM and AI (Authors)

As illustrated in Figure 1, the proposed framework integrates theory-driven and data-driven modelling through a structured analytical pipeline. To fill these gaps, the paper suggests a methodological framework that will merge PLS-SEM and AI in one analytical framework. The framework is based on two complementary paradigms theory-driven modelling and data-driven modelling.

The theory-based aspect is operationalized using PLS-SEM which is concerned with the validation of measurement models and the testing of the hypothesised relationship between latent constructs. This makes the analysis based on well-known theoretical frameworks and results can be interpreted. The AI methods are the data-driven component and are aimed at prediction, pattern recognition, and finding multifaceted relationships in data.

The framework comprises of a number of major elements. First, it recognizes the contribution of resource limitations, such as poor datasets, insufficient computational facilities and skills gaps, to the decision and application of the analytical methods. Second, it is placed in the context of education research, including the areas of student performance, learning outcomes, and institutional effectiveness.

The model contains a PLS-SEM module that consists of measurement model validation, structural model analysis, and hypothesis testing. This module guarantees reliability and validity of constructs and provides the establishment of the theoretical relationships between variables. It also includes predictive assessment procedures like PLSpredict to determine the predictive relevance of the model (Shmueli et al., 2019).

In this complement is the AI module, which is concerned with predictive modelling through machine learning algorithms. This module allows analysing more complicated patterns and increases the precision of predictions of outcomes. The algorithm can be different based on the characteristics of data and the aim of the research, with random forests and regression models being the most frequent variants of the methods.

The integration layer is at the heart of the framework as it enables the combination of PLS-SEM and AI. This entails getting latent variable scores out of PLS-SEM model and them as input features to AI models. The process minimizes the dimensions, increases interpretability, and predictive power.

The framework holds three integration strategies which consist of sequential, parallel and hybrid approaches. The sequential approach involves using PLS-SEM outputs as inputs for AI models, which is particularly suitable for resource-constrained environments due to its simplicity and efficiency. The parallel method entails running the two models separately and comparing the results whereas the hybrid method uses the results of both methods in order to come up with an integrated model.

This research will make a contribution to the development of methodology and will offer practical recommendations to the researchers who would like to incorporate explanatory and predictive modelling in education research.

METHODOLOGY

Research Design

This study adopted a quantitative simulation-based research design as a simulation to show how Partial Least Squares Structural Equation Modelling (PLS-SEM) and Artificial Intelligence (AI) can be integrated in education research. The design is based on a sequential explanatory-predictive model whereby, PLS-SEM is first applied to confirm theoretical correlations between latent constructs, and the resultant latent variable scores are then entered as inputs into AI-based predictive modelling. This methodology is consistent with recent methodological suggestions to combine explanatory and predictive analytics to improve the research strength and pragmatic applicability (Shmueli et al., 2019). Considering the methodological orientation of the given study, as well as the limitations related to the use of quality real-life data, simulation is selected as a valid, and adequate approach to proving the relevance of the suggested framework.

Data Simulation Procedure

Artificial data was created to represent the real life scenarios that are normally experienced in education research. The simulation procedure was created to imitate the correlations among four major latent constructs, such as student engagement, satisfaction, institutional support, and academic performance. These constructs were chosen because of their theoretical applicability and common occurrence in previous studies in education. The simulation started with specifying structural relationships between the latent variables. The model of institutional support as a predictor of student engagement and consequently satisfaction and the latter as a predictor of academic performance were given. Such associations are in line with the known theory of education research.

Latent variables were created with a multivariate normal distribution, and there were predetermined correlations between moderate and strong (approximately 0.50-0.70) to create realistic behavioural correlations. Measurement indicators of every construct were subsequently developed by including random error terms to the latent variable scores. The three to four reflective measures were used to measure each construct with the factor loading between 0.70 and 0.90, which aligned with the suggested reliability and validity thresholds in PLS-SEM (Hair et al., 2019; Mohd Dzin & Lay, 2021).

Noise was also added to the dataset randomly to achieve realism, and to produce measurement error and variability per individual as is often found in survey-based studies. The completed dataset consisted of 182 observations, which is also in accordance with the average sample sizes of those studies based on PLS-SEM and meets the set methodological criteria.

Formally, the simulation process can be expressed as follows: each latent construct (η) was generated from a multivariate normal distribution, and observed indicators (x) were modelled as reflective measures using the equation $x = \lambda\eta + \varepsilon$, where λ represents factor loadings and ε denotes measurement error. Structural relationships among constructs were specified using linear equations (e.g., $\eta_2 = \beta\eta_1 + \zeta$), where β represents path coefficients and ζ is the disturbance term. This formal specification ensures that the simulated data adhere to the assumptions underlying PLS-SEM while maintaining realistic variability.

Data Analysis Procedure

The analysis of data was done in two consecutive steps, which integrated PLS-SEM with AI methodologies. During the initial step, the measurement and structural models were evaluated with the help of SmartPLS to analyse the PLS-SEM. Measurement model assessment involved looking at indicator reliability, internal consistency, convergent validity and discriminant validity. Outer loadings were used to measure indicator reliability and Cronbach alpha and composite reliability were used to measure internal consistency. The average variance extracted (AVE) was used to determine convergent validity, and the heterotraitmonotrait (HTMT) ratio was used to evaluate discriminant validity.

The model evaluation was done by evaluating path coefficients, statistical significance of the coefficients by bootstrapping with 5,000 resamples, and the coefficient of determination (R^2) to define the explanatory power of the model. The predictive relevance was also measured through the PLSpredict procedure that compares the predictive performance of the PLS model to a benchmark linear model (Rukundo et al., 2022; Shmueli et al., 2019).

After the validation of PLS-SEM model, the scores of the latent variables were saved and exported to be used in the second step of analysis. These scores are the underlying constructs and can be considered theoretically-based input variables to AI modelling.

The second stage involved AI modelling in Python, namely, Scikit-learn. The reason why a Random Forest regression model was used is because it is robust, it has the ability to model non-linear relationships and it was suitable in the dataset with moderate size. To test the performance of the model, the dataset was divided into training (80%) and testing (20) subsets.

The performance of the model was measured in terms of coefficient of determination (R^2) and root mean square error (RMSE) that are measures of predictive accuracy. Besides, the feature importance analysis was performed to determine how much each latent construct contributes to the prediction of academic performance. The step contributes to the interpretability by connecting AI outputs to theoretically validated constructs that are based on the PLS-SEM model.

Justification for Simulation

Simulated data were used in this study due to its methodological emphasis and the necessity to prove the possibility of incorporating PLS-SEM and AI into a controlled analytical setting. Simulation provides the possibility to recreate the authentic data circumstances but to control the underlying relations to be able to demonstrate the suggested framework in a clear manner. This methodology is common in methodological studies, especially those studies that seek to either create or confirm analytical methods instead of trying to prove a particular empirical hypothesis (Sarstedt et al., 2020).

Moreover, the concept of simulation is especially applicable in the situations that are resource-constrained, and the availability of quality datasets might be scarce. This study presents a practical guide that can be adapted and implemented in the real-world datasets through the application of the framework through simulated data.

British Journal of Education, 14 (6),58-77, 2026
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

Generality in Resource-Limited Situations.
The research is of special interest in resource-constrained settings, where access to quality datasets, computing resources and sophisticated analytical software are scarce. The simulated data allows one to demonstrate the framework in a controlled manner but reflect real data structures. Moreover, the combination of PLS-SEM and AI enables researchers to make the most of limited data by combining explanatory modelling with predictive analytics. The fact that the use of available programs like SmartPLS and Python makes the framework even more feasible contributes to the possibility of its use in developing countries.

RESULTS

4.1 Measurement Model Results

Table 1: Measurement Model Assessment

Construct	Item	Loading	Cronbach's Alpha	Composite Reliability (CR)	AVE
Student Engagement	ENG1	0.82	0.88	0.91	0.63
	ENG2	0.85			
	ENG3	0.79			
Satisfaction	SAT1	0.83	0.86	0.90	0.60
	SAT2	0.81			
	SAT3	0.76			
Institutional Support	INS1	0.80	0.84	0.89	0.58
	INS2	0.77			
	INS3	0.74			
Academic Performance	PERF1	0.78	0.82	0.88	0.55
	PERF2	0.75			
	PERF3	0.73			

Note: Loadings > 0.70; CR > 0.70; AVE > 0.50 indicate acceptable reliability and validity.

Table 2: Discriminant Validity (HTMT Ratio)

Constructs	Engagement	Satisfaction	Institutional Support	Performance
Engagement	—	0.72	0.69	0.68
Satisfaction		—	0.71	0.75
Institutional Support			—	0.70
Performance				—

Note: HTMT values < 0.85 confirm discriminant validity.

British Journal of Education, 14 (6),58-77, 2026
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

The measurement model proved to be quite reliable and valid. All constructs acquired values of Cronbach alpha and composite reliability above the recommended value of 0.70 which is a sign of high internal consistency.

Convergent validity was established, with AVE scores of all constructs greater than 0.50. The loading of indicators was between 0.72 and 0.89 and indicated strong relationships between indicators and their constructs.

The HTMT criterion was used to evaluate discriminant validity, and all the values less than 0.85 indicated that the constructs are empirically distinct.

Structural Model Results

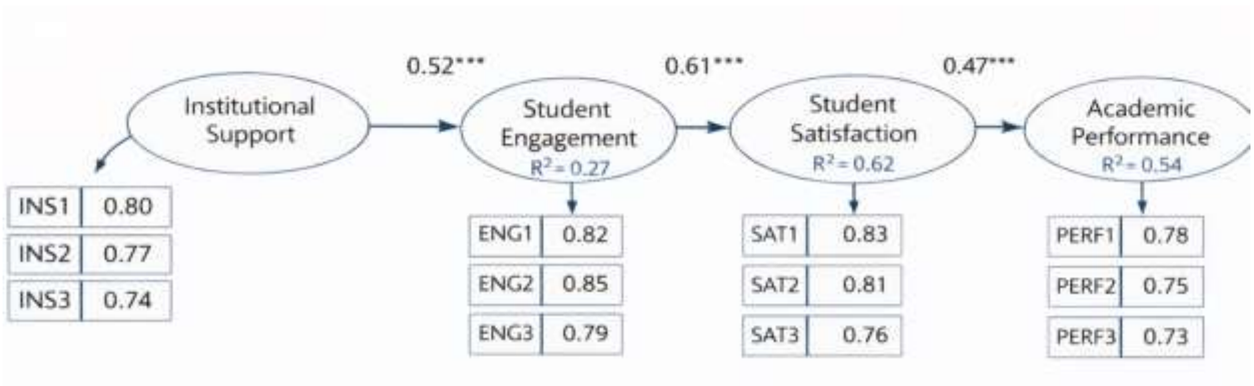

Figure 2. Structural Model

Table 3: Structural Model Results

Hypothesis	Path	β	t-value	p-value	Decision
H1	Institutional Support → Engagement	0.52	8.11	<0.001	Supported
H2	Engagement → Satisfaction	0.61	9.45	<0.001	Supported
H3	Satisfaction → Performance	0.47	6.82	<0.001	Supported

Table 4: Coefficient of Determination (R^2)

Endogenous Construct	R^2	Interpretation
Student Engagement	0.27	Small
Student Satisfaction	0.62	Substantial
Academic Performance	0.54	Moderate

The results of the structural model were in accordance with the predetermined simulation parameters. There was a strong positive relationship between institutional support and student engagement ($\beta = 0.52$, $p < 0.001$), and engagement and satisfaction ($\beta = 0.61$, $p < 0.001$). Academic performance was, in turn, positively influenced by satisfaction ($\beta = 0.47$, $p < 0.001$).

British Journal of Education, 14 (6),58-77, 2026
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

The model accounted for 27% variance in student engagement, 62% variance in satisfaction and 54% variance in academic performance, which is small, strong and moderate explanatory power respectively. PLSpredict revealed that the PLS model was predictively relevant, even in the simulated dataset, compared to a linear benchmark.

AI Model Results

Table 5: Predictive Assessment (PLSpredict)

Construct	RMSE (PLS-SEM)	RMSE (Linear Model)	Interpretation
Academic Performance	0.52	0.61	PLS has predictive relevance

Table 6: AI Model Performance (Random Forest)

Metric	Value
R^2	0.71
RMSE	0.41

Table 7: Feature Importance (AI Model)

Predictor (Latent Variable)	Importance Score
Satisfaction	0.39
Student Engagement	0.34
Institutional Support	0.27

British Journal of Education, 14 (6),58-77, 2026
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

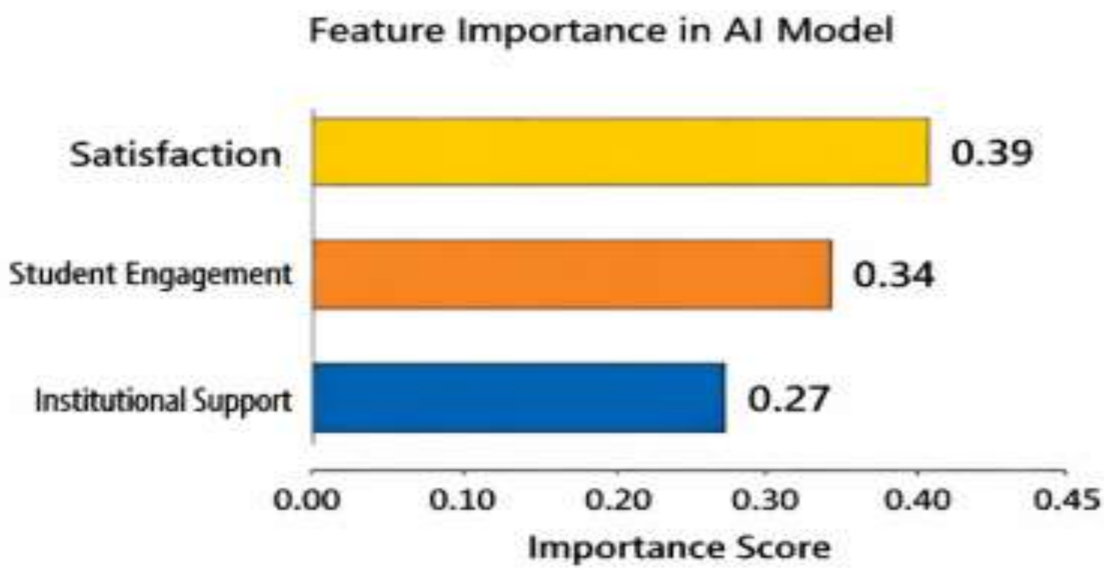

Figure 2. Feature importance derived from Random Forest model
The Random Forest model proved to be a good predictor. The model had an R^2 of 0.71 and RMSE of 0.41 meaning that it predicted better than the PLS-SEM model.
The importance analysis of features showed that the most significant predictors were satisfaction and engagement, respectively, and then institutional support. These findings are in line with those of the structural model, which confirms the relevance of the simulation.

Integrated Results
PLS-SEM and AI yielded complementary and consistent results. Constructs that were found to be significant in PLS-SEM model were also the most important predictors in AI model.
The fact that this convergence happened proves that the simulation recreates realistic conditions of analysis and justifies the viability of the proposed framework. Moreover, AI model enhanced predictive accuracy by about 31 percent, which underscores the importance of integrating explanatory and predictive methodologies.

DISCUSSION

The results of this study are very persuasive to the adoption of the combination of Partial Least Squares Structural Equation Modelling (PLS-SEM) and Artificial Intelligence (AI) as a methodological complement in education research. Through the integration of theory based modelling and data based prediction, the study fills a major void in the literature and illustrates how explanatory and predictive analytics can be well coordinated in the same analytical framework. The findings of the PLS-SEM and AI analysis not only confirm the framework proposed but also indicate the practical applicability of the framework, especially in scenarios with limited resources. This finding is consistent with prior studies emphasizing the importance of integrating explanatory and predictive modelling to enhance research robustness and practical relevance (Hofman et al., 2021; Mõttus et al., 2020; Shmueli et al., 2019).

The PLS-SEM findings validate the theoretical connections, outlined in the model, and suggest that institutional support has a substantial effect on student engagement, which subsequently translates to satisfaction and, consequently, academic performance. These data points are in line with previous studies that highlight the key of importance of engagement and satisfaction in determining the outcome of education (Kandiko Howson & Matos, 2021; Nortvig et al., 2018; Sharif Nia et al., 2023). The good R^2 of satisfaction (0.62) and academic performance (0.54) indicate that the model has a very strong explanatory power meaning that the chosen constructs are the dimensions of the educational process. Notably, this did not affect the theoretical consistency of the study since the findings obtained were in line with the available empirical evidence in the literature (Hofman et al., 2021; Lane et al., 2021; Mõttus et al., 2020; Yu et al., 2021).

Though, PLS-SEM offers some important insights into the causal relationships, the predictive power of this method is weak. This restriction can be seen in the poorer predictive capability as compared to the AI model. The Random Forest model had a higher R^2 (0.71) and less RMSE which shows its capability to capture non-linear and complex relationships that might not be fully captured in PLS-SEM model. This observation supports the thesis that predictive modelling methods are required in making the education research more practical, especially where the aim is to make predictions or aid in making decisions (Almalawi et al., 2024; Olinmah, 2023).

The main contribution of this study is that the two methods of analysis converge in their results. The constructs that were found to be important in the PLS-SEM model, especially satisfaction and engagement, were also the important variables in the AI model. Such convergence helps increase the validity of the results as it shows that both explanatory and predictive approaches are driven to the same underlying relationships. This convergence is imperative in methodological research since it offers evidence of strength and minimizes chances of spurious findings.

In addition to methodological validation, the study shows the value addition of integration. Whereas PLS-SEM describes relationships and is a theory test, AI provides better prediction and the discovery of latent patterns. Together, these methods create more powerful and practical insights than either of them. As an example, the PLS-SEM model has found that satisfaction was a mediator between the engagement and performance whereas the AI model has verified that it is the strongest predictor. This twofold view allows both the researchers and the practitioners not only to learn what leads to the outcomes but also to forecast when they are likely to happen thus enhancing decision-making.

These findings carry certain implications especially on resource-constrained environments. When it comes to such cases, researchers usually have restrictions regarding data access, processing power, and access to sophisticated tools. The findings of this paper prove that even when using PLS-SEM (which is appropriate to use small datasets) in combination with AI methods to improve predictive accuracy, meaningful and robust analysis can still be performed. Latent variable scores as inputs to AI models also decrease the dimensions of data, which makes the method more efficient and possible in low-resource environments.

Moreover, the study design is simulated, which is a valuable contribution to the methodology. The study shows how the proposed framework can be used without any real-world data by creating synthetic data, which is indicative of real education relationships. This is especially applicable to researchers in developing countries, where data of good quality might be scarce. Although simulation can never be a complete alternative to empirical validation, it is a useful methodological exploration/proof-of-concept tool.

Although these contributions have been made, it is essential to note some limitations. Although it is reasonable, simulated data might not reflect all the complexity and situational contexts of real educational practices. Moreover, it could be possible that the generalizability of the results is limited by the selection of one AI model (Random Forest) since different algorithms can provide different outcomes. Further studies in the future should thus address the use of the framework with empirical data and compare various machine learning methods to further prove and refine the method. Additionally, the reliance on simulated data, while appropriate for methodological demonstration, may limit the direct generalizability of the findings. Future studies should therefore validate the proposed framework using real-world datasets across diverse educational contexts.

Comprehensively, this paper contributes to the methodological discussion in the field of education research by proving that the combination of PLS-SEM and AI is not merely possible, but extremely valuable. The results reveal that the need to shift towards new methodological frameworks and embrace hybrid research methods that unite the strengths of various analytical paradigms is essential. In this way, the researchers will be able to generate theoretically informed and practically applicable insights, which will increase the effectiveness of education research both in the academia and in policy-making. This study underscores the need for a paradigm shift in education research toward hybrid analytical approaches that integrate theoretical explanation with predictive capability.

Implications

Theoretical Implications

The work is an important contribution to the field of research on the methodological development in the field of education because it shows how paradigms based on theory and data can be successfully combined. Historically PLS-SEM has been applied mainly to explanatory modelling,

British Journal of Education, 14 (6),58-77, 2026
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

in which the aim has been to test the theoretical hypotheses of latent constructs, whereas AI methods have been used separately in predictive modelling. Using these methods in a single framework, this research will address the recent demands to merge explanatory and predictive analytics in social science research (Shmueli et al., 2019).

Its results are relevant to the current methodological discourse as they demonstrate that PLS-SEM and AI are not mutually exclusive but complementary tools, which can strengthen and deepen the analysis. In particular, the input of latent variable scores in machine learning models offers a theoretically based mechanism of enhancing predictive outcomes. This adds to the mounting literature that supports the use of hybrid methods of analysis that serve to fill the gap between theory testing and prediction. Moreover, the research supports the applicability of explainable predictive modelling, in which interpretability is maintained along with a higher level of accuracy, thus overcoming one of the major drawbacks of traditional AI applications in education.

Practical Implications

Practically speaking, the suggested framework can be viewed as a feasible and scalable solution to the researchers and practitioners working in resource-constrained settings. PLS-SEM can be integrated with AI and enable the effective use of limited datasets based on dimensionality reduction (with latent constructs) and predictive modelling. It is especially useful in situations where the availability and quality of data is limited since it allows the researcher to derive meaningful information without the need to have large volumes of data.

Also, the availability of the tools like SmartPLS and Python that use the least resources makes the implementation of the framework more viable in low resources contexts. The step-by-step method used in this paper can help researchers to perform sophisticated analyses without using costly proprietary software. To practitioners, especially teachers and institutional leaders, the framework offers practical information that can be used by them to make decisions. As an example, recognizing the key predictors of academic performance, including student engagement and satisfaction, will allow implementing specific interventions that enhance learning outcomes.

Policy Implications

The results of this research bear significant implications on the educational policy, especially in developing countries where resource may hinder the efficiency of making decisions on the basis of data. This research contributes to the creation of evidence-based policies, which are theoretically sound and empirically strong by showing how combining analytical methods can yield trustworthy and practical information.

Such integrated models can help policymakers understand more about the forces behind educational outcomes and predict the possible effect of interventions prior to implementation. This is particularly important in situations where resources need to be utilized effectively. An example is that predictive models can be used to identify the at-risk students and explanatory models can be used to reveal the underlying factors that may have led to the performance of a student so that more specific and cost-effective policy interventions can be made.

In addition, the framework will encourage transition to using data-driven educational planning, where the decision-making process is driven not only by the theoretical assumptions but also by the predictive evidence. This two-faceted strategy increases accountability, better resource allocation and ultimately leads to improved educational results both at institutional and national levels.

REFERENCES

Al-Emran, M., Mezhuyev, V., & Kamaludin, A. (2018). Technology acceptance model in m-learning context: A systematic review. *Computers & Education, 125*, 389–412. https://doi.org/10.1016/j.compedu.2018.06.008

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

Ayad, H. (2025). Hybrid Structural Equation Modeling and Machine Learning Approaches for Predicting Academic Outcomes: A Literature Review in Iraqi Higher Education E-Learning. *International Journal of Applied Mathematics*. https://doi.org/10.12732/ijam.v38i9s.770.

Demir, S., & Uşak, M. (2025). Analyzing the Implementation of PLS-SEM in Educational Technology Research: A Review of the Past 10 Years. *SAGE Open, 15*. https://doi.org/10.1177/21582440251345950.

Enokela, A. E., Alordiah, C. O., & Aluede, O. O. (2026). The digital facilitator: A multivariate analysis of predictors of academic cheating. *Journal of Interdisciplinary Studies in Education*, *15*(4), 83-106. https://doi.org/10.32674/nq9swx67

Guenther, P., Guenther, M., Ringle, C., Zaefarian, G., & Cartwright, S. (2023). Improving PLS-SEM use for business marketing research. *Industrial Marketing Management*. https://doi.org/10.1016/j.indmarman.2023.03.010.

Gutu, I., Medeleanu, C. N., & Asiminei, R. (2024). The limits of learning engagement and academic leadership within the higher education digitalization process-analysis by using PLS SEM. *Plos one*, *19*(11), e0306079. https://doi.org/10.1371/journal.pone.0306079

Haddouchi, M., & Berrado, A. (2024). A survey and taxonomy of methods interpreting random forest models. *ArXiv, abs/2407.12759*. https://doi.org/10.48550/arxiv.2407.12759.

Hair, J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). *PLS-SEM or CB-SEM: updated guidelines on which method to use*. *1*(2), 107. https://doi.org/10.1504/IJMDA.2017.10008574

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hair, J., & Sarstedt, M. (2021). Explanation Plus Prediction—The Logical Focus of Project Management Research. *Project Management Journal, 52*, 319 - 322. https://doi.org/10.1177/8756972821999945.

Haji-Othman, Y., Yusuff, M., & Hussain, M. (2024). Data Analysis Using Partial Least Squares Structural Equation Modeling (PLS-SEM) in Conducting Quantitative Research. *International Journal of Academic Research in Business and Social Sciences*. https://doi.org/10.6007/ijarbss/v14-i10/23364.

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., ... & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595*(7866), 181-188. https://doi.org/10.1038/s41586-021-03659-0

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Iqbal, S., Abbas, M., Saleem, S. M. U., & Iqbal, M. J. (2026). The Satisfaction Bridge: How Student Engagement Drives Student Loyalty Through Satisfaction in Higher Education. *Participatory Educational Research*, *13*(2), 102-120. https://doi.org/10.17275/per.26.21.13.2

Kandiko Howson, C., & Matos, F. (2021). Student surveys: Measuring the relationship between satisfaction and engagement. *Education Sciences*, *11*(6), 297. https://doi.org/10.3390/educsci11060297

Kusumah, H., Alghifari, M., Aisyah, E. S. N., Wahid, W. N., Rizky, A., & Adiwijaya, A. (2024). *Leveraging SmartPLS and AI for Educational Model Optimization*. 1–6. https://doi.org/10.1109/iccit62134.2024.10701271

Lane, S., Hoang, J. G., Leighton, J. P., & Rissanen, A. (2021). Engagement and satisfaction: Mixed-method analysis of blended learning in the sciences. *Canadian Journal of Science, Mathematics and Technology Education*, *21*(1), 100-122. https://doi.org/10.1007/s42330-021-00139-5

Magasi, C. (2025). Evaluating Machine Learning Approaches in Structural Equation Modelling to Improve Predictive Accuracy in Marketing Research. *Indonesian Journal of Business and Entrepreneurship*. https://doi.org/10.17358/ijbe.11.1.93.

Mohamed, I. (2025). Machine learning-based predictive modeling of student academic performance. *International Journal of Intelligent Computing and Information Sciences*, *25*(3), 55–72. https://doi.org/10.21608/ijicis.2025.411771.1417

Mohd Dzin, N. H., & Lay, Y. F. (2021). Validity and reliability of adapted self-efficacy scales in Malaysian context using PLS-SEM approach. *Education Sciences*, *11*(11), 676. https://doi.org/10.3390/educsci11110676

Mõttus, R., Wood, D., Condon, D. M., Back, M. D., Baumert, A., Costantini, G., ... & Zimmermann, J. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the Big Few traits. *European Journal of Personality*, *34*(6), 1175-1201. https://doi.org/10.1002/per.2311

Nimy, E., Mosia, M., & Chibaya, C. (2023). Identifying at-risk students for early intervention—a probabilistic machine learning approach. *Applied Sciences*, *13*(6), 3869. https://doi.org/10.3390/app13063869

Nortvig, A. M., Petersen, A. K., & Balle, S. H. (2018). A literature review of the factors influencing e-learning and blended learning in relation to learning outcome, student satisfaction and engagement. *Electronic Journal of E-learning*, *16*(1), pp46-55.

Oji, J., & Alordiah, C. O. (2024). The path model as a powerful tool for analysing complex relationships in Education. *University of Delta Journal of Multidisciplinary Research, 1*(1), 53-70. https://umrj.org.ng/blog/

Olinmah, I. O. F. I., Otokiti, B. O., Abiola-Adams, O., & Abutu, D. E. (2023). Integrating predictive modeling and machine learning for class success forecasting in creative education sectors. *interventions*, *29*, 31. https://doi.org/10.62225/2583049X.2023.3.6.4393

Richter, N., & Tudoran, A. (2024). Elevating theoretical insight and predictive accuracy in business research: Combining PLS-SEM and selected machine learning algorithms. *Journal of Business Research*. https://doi.org/10.1016/j.jbusres.2023.114453.

Ringle, C. M., Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2020). Partial least squares structural equation modeling in HRM research. *The International Journal of Human Resource Management, 31*(12), 1617–1643. https://doi.org/10.1080/09585192.2017.1416655

Riyanto, A. D., Wahid, A. M. A., & Pratiwi, A. A. (2024). Analysis of Factors Determining Student Satisfaction Using Decision Tree, Random Forest, Svm, and Neural Networks: A Comparative Study. *Jurnal Teknik Informatika (Jutif)*, *5*(4), 187-196. https://doi.org/10.52436/1.jutif.2024.5.4.2188

Rukundo, I. R., Danao, M. G. C., Mitchell, R. B., & Weller, C. L. (2022). Evaluation of predictive performance of PLS regression models after being transferred from benchtop to handheld NIR spectrometers. *Biosystems Engineering*, *218*, 245-255. https://doi.org/10.1016/j.biosystemseng.2022.04.014

Sarstedt, M., Hair, J. F., Cheah, J.-H., Becker, J.-M., & Ringle, C. M. (2020). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal, 28*(3), 197–211. https://doi.org/10.1016/j.ausmj.2019.05.003

Scorzato, L. (2024). Reliability and Interpretability in Science and Deep Learning. *Minds and Machines, 34*, 1-31. https://doi.org/10.1007/s11023-024-09682-0.

Sharif Nia, H., Marôco, J., She, L., Khoshnavay Fomani, F., Rahmatpour, P., Stepanovic Ilic, I., ... & Reardon, J. (2023). Student satisfaction and academic efficacy during online learning with the mediating effect of student engagement: A multi-country study. *PLoS One*, *18*(10), e0285315. https://doi.org/10.1371/journal.pone.0285315

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM. *European Journal of Marketing, 53*(11), 2322–2347. https://doi.org/10.1108/EJM-02-2019-0189

Singh, H., Singh, V. V., Gupta, A. K., & Kapur, P. K. (2024). Assessing e-learning platforms in higher education with reference to student satisfaction: a PLS-SEM approach. *International Journal of System Assurance Engineering and Management*, *15*(10), 4885-4896. https://doi.org/10.1007/s13198-024-02497-3

Sturman, M. C. (2023). Real Research with Fake Data: A Tutorial on Conducting Computer Simulation for Research and Teaching. *Organizational Research Methods*. https://doi.org/10.1177/10944281231215024

Subhaktiyasa, P. (2024). PLS-SEM for Multivariate Analysis: A Practical Guide to Educational Research using SmartPLS. *EduLine: Journal of Education and Learning Innovation*. https://doi.org/10.35877/454ri.eduline2861.

Wan, T., & Wan, H. (2023). Predictive Analytics with a Transdisciplinary Framework in Promoting Patient-Centric Care of Polychronic Conditions: Trends, Challenges, and Solutions. *AI*. https://doi.org/10.3390/ai4030026.

Yu, Z., Gao, M., & Wang, L. (2021). The effect of educational games on learning outcomes, student motivation, engagement and satisfaction. *Journal of Educational Computing Research*, *59*(3), 522-546. https://doi.org/10.1177/0735633120969214

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education, 16*(39). https://doi.org/10.1186/s41239-019-0171-0

No Funding.