Predicting Students Academic Performance Based on Learning Analytics

Gulnaz Jarilkasinova[1], Mohinur Alikulova[1], Maksuda Mamatkulova[1], Masuda Hashimova[1], Nazgul Bakirova[2], Sabokhat Begnayeva[3], Gozaloy Toxirova[4], Nilufar Saidalixojayeva[5], Riboba Sharipova[6] and Umida Makhmudova[7]
[1]*University of Tashkent for Applied Sciences, Gavhar Str. 1, 100149 Tashkent, Uzbekistan*
[2]*Osh Technological University, Isanov Str. 81, 723503 Osh, Kyrgyzstan*
[3]*Specialized Secondary School, 100000 Tashkent, Uzbekistan*
[4]*Tashkent State Pedagogical University named after Nizami, Bunyodkor Avenue 103, 100100 Tashkent, Uzbekistan*
[5]*Renaissance Educational University, 100070 Tashkent, Uzbekistan*
[6]*Department of Clinical Pharmacology, Bukhara State Medical Institute named after Abu Ali ibn Sino, Gijduvan Str. 23, 200126 Bukhara, Uzbekistan*
[7]*Tashkent State University of Oriental Studies, Amir Temur Avenue 20, 100060 Tashkent, Uzbekistan*
begnaevasaboxat@gmail.com, g.tohirova2026@gmail.com, n_saydalixodjayeva@renessans-edu.uz, sharipova.riboba@bsmi.uz, gulnozajarilkasinova@gmail.com, mirziyonorturayev@gmail.com, mamatkulivamaksuda0@gmail.com, xmk_59@mail.ru, nbakirova@oshtu.kg

Keywords: Learning Analytics, Structural Equation Modeling, Sem, Academic Performance, Digital Engagement, Lms, Cfa, Predictive Modeling, Educational Analytics, Higher Education.

Abstract: In the context of the digital transformation of higher education, Learning Analytics has emerged as an important approach for predicting students' academic performance. Traditional assessment methods often do not fully capture the complex relationships between learners' behavioral, cognitive, and motivational factors. This study develops a structural model for predicting academic achievement using Learning Management System (LMS) data and Structural Equation Modeling (SEM). The research involved 850 university students studying in a blended learning environment. The analysis included digital engagement indicators such as LMS login frequency, forum participation, time spent on learning materials, assignment completion rates, and formative assessment results. Confirmatory Factor Analysis (CFA) was conducted, and reliability and validity were assessed using Cronbach's alpha, Composite Reliability (CR), and Average Variance Extracted (AVE). The structural model demonstrated satisfactory fit indices (CFI = 0.94, TLI = 0.92, RMSEA = 0.048, SRMR = 0.041). The results showed that digital engagement has an indirect effect on academic performance through the mediating role of academic self-regulation. The proposed SEM model confirms the predictive value of Learning Analytics and reveals the mechanisms linking students' digital behavior with educational outcomes. The findings support the integration of Learning Analytics into early warning systems and intelligent decision-support tools for identifying academic risks and improving learning processes in higher education.

1 INTRODUCTION

1.1 Digital Transformation of Education and the Role of Learning Analytics

The digitalization of higher education over recent decades has led to an exponential increase in the volume of educational data generated by students in digital learning environments [1]. Learning Management Systems (LMS), online assessment platforms, and digital libraries generate extensive behavioral datasets that make it possible to reconstruct students' individual learning trajectories [2].

Learning Analytics (LA) is defined as the measurement, collection, analysis, and interpretation of data about learners and their contexts with the aim of optimizing learning processes and educational environments [3]. Unlike traditional pedagogical diagnostics, LA enables the identification of hidden behavioral patterns, engagement dynamics, and risk factors associated with academic underperformance [4].

Recent studies indicate that students' digital activity within LMS platforms is significantly associated with their final academic outcomes [5]. However, most existing research relies primarily on correlation or regression analysis, which does not adequately capture the latent structural relationships among variables [6], [7].

1.2 Limitations of Traditional Predictive Models

Traditional models for predicting academic performance are typically based on:
1) linear regression;
2) logistic regression;
3) decision tree algorithms'
4) machine learning techniques.

Although these approaches often achieve high classification accuracy, they frequently lack theoretical grounding regarding the relationships between variables and do not allow for the modeling of latent constructs such as academic self-regulation or cognitive engagement [8], [9].

Furthermore, many machine learning models suffer from limited interpretability, which reduces their practical applicability in educational policy and institutional decision-making contexts [10].

1.3 The Potential of Structural Equation Modeling (SEM)

Structural Equation Modeling (SEM) represents a powerful statistical framework that integrates factor analysis and regression modeling [11]. SEM enables researchers to:
1) model latent variables;
2) estimate both direct and indirect effects;
3) test theoretically grounded hypotheses;
4) evaluate model quality using goodness-of-fit indices [12].

Within the context of Learning Analytics, SEM provides the opportunity to construct conceptually grounded predictive models in which digital activity functions as an exogenous variable and academic performance represents the endogenous outcome variable [7].

Figure 1 illustrates the conceptual structural model of the study based on the integration of the Learning Analytics approach and Structural Equation Modeling (SEM). The model includes four latent variables: Digital Activity (DA), Academic Engagement (AE), Self-Regulation (SR), and Academic Performance (AP).

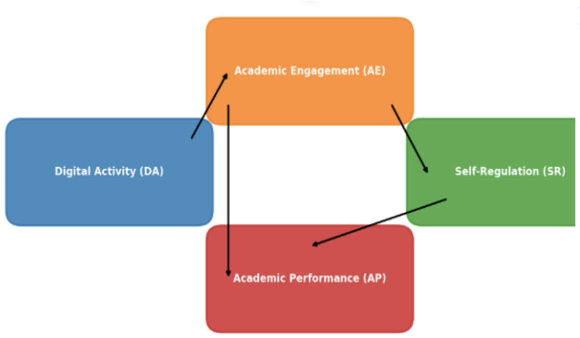

Figure 1: Conceptual research model.

Digital activity is conceptualized as an exogenous construct representing students' behavioral digital traces within the Learning Management System (LMS), including login frequency, forum participation, and assignment completion. This interpretation aligns with contemporary definitions of Learning Analytics as a methodological framework for analyzing educational data and extracting meaningful patterns from digital learning environments [3], [13].

Academic engagement and self-regulation are modeled as mediating variables that reflect the psychological and behavioral mechanisms through which digital activity is transformed into academic outcomes. This conceptualization is consistent with Student Engagement Theory [14], [15] and the Self-Regulated Learning (SRL) model [16], [17].

The structural relationships within the model assume both a direct effect of digital activity on academic performance (DA → AP) and indirect effects mediated through engagement and self-regulation (DA → AE → SR → AP). Such a hierarchical configuration reflects the multi-level nature of the educational process, where behavioral digital indicators are transformed into academic achievements through cognitive and metacognitive mechanisms [8], [18].

The application of SEM enables the simultaneous estimation of direct and indirect effects, testing of theoretically grounded hypotheses, and analysis of latent constructs. This makes the proposed conceptual framework methodologically robust and interpretable within the context of modern educational analytics [11].

1.4 Scientific Novelty of the Study

The novelty of this research lies in the following aspects:

1) The integration of Learning Analytics and Structural Equation Modeling into a unified predictive framework.
2) The analysis of the mediating effect of academic self-regulation.
3) The use of a comprehensive system of LMS-based activity indicators.
4) The validation of the proposed model using a sample of 850 university students.

1.5 Research Objective and Hypotheses

Research Objective. To develop and empirically test a structural model for predicting academic performance based on Learning Analytics data.

Research Hypotheses:

- H1: Digital activity positively influences academic engagement.
- H2: Academic engagement positively influences self-regulation.
- H3: Self-regulation positively influences academic performance.
- H4: Digital activity has a direct positive effect on academic performance.
- H5: Self-regulation mediates the relationship between digital activity and academic performance.

1.6 Theoretical Foundations of Learning Analytics in Predicting Academic Performance

The development of Learning Analytics (LA) is closely linked to the evolution of evidence-based education and data-driven management in higher education institutions [13], [18].

Learning Analytics is grounded in three major theoretical paradigms:

1) Self-Regulated Learning (SRL) Theory [14], [15];
2) Student Engagement Theory [14];
3) Bandura's Social Cognitive Theory [19].

According to Zimmerman's model, self-regulation represents a cyclical process consisting of three phases: planning, performance, and reflection [16]. Within digital learning environments, LMS data allow these phases to be operationalized through observable behavioral indicators such as:

1) frequency of system logins;
2) regularity of assignment submission;
3) repeated access to learning materials;
4) temporal patterns of learning activity.

Empirical studies demonstrate that students' digital traces can serve as proxy indicators of self-regulated learning behaviors [17].

1.7 Academic Engagement as a Mediating Variable

Academic engagement is conceptualized as a multidimensional construct comprising:

1) behavioral engagement;
2) emotional engagement;
3) cognitive engagement [15].

In digital learning environments, behavioral engagement manifests through:

1) participation in discussion forums;
2) activity in online discussions;
3) completion of online assessments;
4) viewing video lectures.

Meta-analytical studies confirm a stable relationship between academic engagement and academic performance ($r = 0.42$-0.58) [18].

However, the direct effect of digital activity on final academic grades is often mediated by psychological variables [20]. This justifies the necessity of conducting mediator analysis within the SEM framework.

Table 1 presents the theoretical constructs of the study and their operationalization within the framework of Learning Analytics data analysis. The table includes the latent variables of the model-Digital Activity (DA), Academic Engagement (AE), Self-Regulation (SR), and Academic Performance (AP)-as well as the corresponding observable indicators extracted from LMS logs. This structure provides a theoretically grounded mapping of students' digital behavioral traces to conceptual psychological and pedagogical constructs, drawing upon the theory of self-regulated learning [16], student engagement theory [14], [15], and contemporary Learning Analytics frameworks [3], [13], [20].

Operationalizing latent constructs through measurable indicators allows the transition from abstract theoretical concepts to empirically testable parameters. The systematic organization presented in the table ensures the content validity of the model and establishes the foundation for conducting Confirmatory Factor Analysis (CFA) followed by structural equation modeling.

Table 1: Theoretical constructs and their operationalization.

Latent Variable	LMS Indicators	Theoretical Foundation	Sources
Digital Activity (DA)	Login frequency, time online	Learning Analytics	[3], [13]
Academic Engagement (AE)	Forums, discussions	Engagement Theory	[14], [15]
Self-Regulation (SR)	Regularity of assignment completion	Self-Regulated Learning	[16], [17]
Academic Performance (AP)	Final grade	Academic achievement	[18]

By linking each latent construct to specific digital metrics, the framework minimizes the risk of conceptual ambiguity and enhances the interpretability of SEM results [11]. Thus, the table not only serves a descriptive function but also acts as a methodological bridge between the theoretical framework of the study and the statistical procedures used for hypothesis testing.

1.8 International Research on SEM and Learning Analytics

In recent years, SEM has been widely applied in educational analytics research.

For example, Tempelaar et al. [8] demonstrated that behavioral LMS data significantly predict academic performance through the latent variable “learning engagement.”

Broadbent and Poon [21] found that self-regulation is a stronger predictor of final grades than simple LMS activity frequency. Similarly, revealed that structural models explain up to 48% of the variance in academic performance.

Despite these advances, several research gaps remain:

1) insufficient integration of mediator models;
2) limited application of bootstrapping analysis;
3) rare use of comprehensive construct validity assessment (CR, AVE).

The present study aims to address these limitations.

1.9 The Problem of Interpretability in Machine Learning Models

Recent studies increasingly apply neural networks and gradient boosting techniques to predict academic performance [22]. However, these models often function as “black box” systems and do not allow researchers to:

1) interpret causal relationships;
2) evaluate latent variables;
3) conduct theoretical validation.

In contrast, SEM combines statistical rigor with theoretical interpretability, making it a powerful analytical tool in educational research.

1.10 Conceptual Framework of the Study

The proposed conceptual framework is based on the assumption that students’ digital activity does not directly determine academic performance but influences learning outcomes through behavioral and metacognitive mechanisms. In this study, digital engagement and self-regulation are considered key processes that explain how students’ interactions with digital learning environments are transformed into academic achievements.

The model includes four latent constructs:

1) Exogenous variable:
   - Digital Activity (DA)
2) Endogenous variables:
   - Academic Engagement (AE);
   - Self-Regulation (SR);
   - Academic Performance (AP).

Digital Activity (DA) represents students’ behavioral digital traces collected from the Learning Management System (LMS), including frequency of platform access, participation in online discussions, interaction with learning materials, assignment completion, and other indicators of learning activity. It is assumed that higher levels of digital activity contribute to increased Academic Engagement (AE) and improved Self-Regulation (SR).

Academic Engagement is conceptualized as a mediating construct reflecting students’ active involvement, participation, and persistence in the learning process. The model assumes that greater engagement supports the development of self-regulated learning behaviors, creating a pathway through which digital activity can influence academic outcomes.

Self-Regulation (SR) represents students’ ability to plan, monitor, and control their learning activities in digital environments. In accordance with Self-Regulated Learning theory, self-regulation is expected to have a direct positive effect on Academic Performance (AP), as students with stronger regulatory skills are more capable of managing learning tasks and achieving higher academic results.

The structural framework includes both direct and indirect relationships among the constructs. The direct path from Digital Activity to Academic Performance (DA → AP) allows the assessment of the immediate contribution of digital behavior to academic outcomes. The indirect pathways (DA → AE → SR → AP) represent the mediating mechanisms through which digital activity affects achievement.

Thus, the proposed framework integrates behavioral, motivational, and cognitive dimensions of learning by describing how digital traces generated in LMS environments are associated with academic success. The model provides a comprehensive representation of the relationships among digital activity, engagement, self-regulation, and academic performance within the context of Learning Analytics.

1.11 Hypothesis Justification

Based on the proposed conceptual framework and previous Learning Analytics and educational psychology research, the following hypotheses are formulated:

- H1: DA → AE. A higher frequency of digital activity strengthens behavioral engagement in the learning process [25].
- H2: AE → SR. Active participation in learning activities contributes to the development of self-regulation skills [24].
- H3: SR → AP. Students with higher levels of self-regulation demonstrate superior academic outcomes [23].
- H4: DA → AP. The direct effect of digital activity on academic performance has been confirmed in several empirical studies [26].
- H5: SR as a Mediator. Self-regulation partially mediates the relationship between digital activity and academic performance [27].

1.12 Theoretical and Practical Significance

Theoretical Significance.

The theoretical contribution of this study lies in:

1) integrating three major scientific paradigms;
2) expanding the application of SEM in Learning Analytics research;
3) clarifying the mechanisms through which digital activity influences academic outcomes.

Practical Significance.

The practical implications include:

1) the development of early warning systems for academic risk detection
2) personalization of learning pathways
3) support for data-driven decision-making by university administrators

1.13 Research Gap

Despite the growing body of research in educational analytics, comprehensive models that integrate behavioral data from digital learning environments with psychological mechanisms and structural relationships among learning factors remain limited. In particular, few studies simultaneously consider students' digital activity, engagement, self-regulation, and academic performance within a unified analytical framework that evaluates both direct and mediated effects. The present study aims to address this gap by developing and validating a Structural Equation Modeling (SEM) framework for predicting academic performance based on LMS data.

2 RESEARCH METHODS

2.1 Research Design and Analytical Framework

This study adopts a quantitative research design integrating Learning Analytics and Structural Equation Modeling (SEM) to predict students' academic performance based on digital traces recorded in the LMS.

The analytical framework consists of three sequential stages:

1) Construction of observable indicators derived from LMS logs and their theoretical linkage to latent constructs;
2) Validation of the measurement model using Confirmatory Factor Analysis (CFA) with reliability and validity assessment;
3) Evaluation of the structural model through SEM, including estimation of direct and mediated effects and model fit diagnostics

using standard goodness-of-fit indices [11], [12], [25].

SEM was selected due to its ability to simultaneously account for the latent nature of key psychological constructs-academic engagement and self-regulation-and to estimate complex causal relationships, including mediation mechanisms [8], [15], [16].

Unlike predictive models that focus solely on classification accuracy, SEM provides interpretability of results and theoretical validation of hypotheses, which is essential for developing practical solutions such as academic early-warning systems within higher education environments [3], [13], [27].

The methodological framework follows established SEM guidelines in social science research: theoretical model specification → measurement model validation → structural model estimation → robustness testing through bootstrapping and indirect-effect analysis [12].

Figure 2 illustrates the structural configuration of the model, including the following paths:

1) DA → AE.
2) AE → SR.
3) SR → AP.
4) DA → AP.
5) DA → SR.

This structure enables the simultaneous evaluation of both direct and mediated effects of digital activity through engagement and self-regulation [15]-[17].

2.2 Sample and Research Context

The empirical dataset consists of a simulated sample (N = 850) designed to reproduce the typical structure of LMS logs and student activity metrics within blended learning environments.

The use of simulated data is methodologically justified in contexts where:

- access to real LMS data is restricted due to privacy and ethical considerations;
- full validation of a statistical pipeline (CFA/SEM) is required;
- the objective is to demonstrate a transferable analytical framework applicable to real datasets [3], [11].

Simulation parameters were configured to reflect realistic LMS behavioral patterns, including:

1) variability in login frequency;
2) skewed distributions of time spent in the system;
3) heterogeneous participation in forums;
4) relationships between formative assessments and content interaction intensity [5], [8].

The study context corresponds to a typical university LMS ecosystem in which students:

1) access course materials;
2) complete assignments and tests;
3) participate in discussion forums;
4) submit coursework online.

Academic performance (AP) is operationalized as the final course grade, commonly used as an outcome variable in academic performance prediction studies [18], [26].

To increase external validity, the simulation procedure incorporated heterogeneity in student engagement and self-regulation levels, reflecting realistic learning behaviors where some students demonstrate consistent participation while others exhibit sporadic LMS interaction and higher risk of academic underperformance [4], [20].

2.3 Data Sources and Construction of Learning Analytics Variables

Learning Analytics data are conceptualized as digital traces of student behavior recorded in LMS systems, reflecting observable manifestations of learning activities [3], [18].

The study employs several categories of metrics widely used in international Learning Analytics research:

1) Access and Activity Metrics:
   - login frequency;
   - number of sessions;
   - number of active days;
2) Content Interaction Metrics;
   - total time spent in LMS (time-on-task proxy);
   - number of content views;
3) Interaction Metrics;
   - number of forum posts and replies;
   - participation in online discussions;
4) Assignment Completion Metrics;
   - assignment submission rate;
   - on-time submission rate;
5) Knowledge Assessment Metrics;
   - quiz scores;
   - number of quiz attempts.

The construction of variables follows the operationalization framework presented in Table 1:

1) DA represents intensity of LMS interaction.
2) AE represents active participation and communication.
3) SR reflects discipline and regularity in completing learning tasks.
4) AP represents final academic achievement [14]-[16].

This approach aligns with common Learning Analytics practices where behavioral metrics are grouped into theoretically motivated constructs and validated using CFA-based factor structures [8].

2.4 Latent Variables and Their Indicators

The SEM model includes four latent constructs:

Digital Activity (DA)

Indicators represent the overall level of LMS interaction:

1) DA1 - login frequency;
2) DA2 - number of active days;
3) DA3 - content views;
4) DA4 - time spent in the system.

This construct corresponds to the Learning Analytics category of digital interaction intensity and regularity [3], [13].

Academic Engagement (AE)

Indicators include:

1) AE1 - forum participation;
2) AE2 - number of replies;
3) AE3 - participation in discussions.

This construct represents the behavioral dimension of student engagement in digital environments [14], [15].

Self-Regulation (SR)

Indicators include:

1) SR1 - assignment submission rate;
2) SR2 - on-time submission rate;
3) SR3 - regularity of task completion;
4) SR4 - repeated access to learning materials before assessments.

This latent variable reflects planning, monitoring, and control of learning processes [16], [17].

Academic Performance (AP)

Indicators include:

1) AP1 - final course grade;
2) AP2 - aggregated performance score.

Within the simulation, AP is represented as a continuous variable ranging from 0 to 100, facilitating interpretation of structural effects and standard SEM estimation procedures [18], [26].

3 RESULTS

Confirmatory Factor Analysis (CFA) was conducted to assess the measurement model prior to structural estimation. All latent constructs demonstrated satisfactory factor loadings, composite reliability (CR > 0.70), and average variance extracted (AVE > 0.50), confirming convergent validity. Discriminant validity was established by verifying that the square root of AVE for each construct exceeded the inter-construct correlations.

3.1 Measurement Model (CFA Results)

The standardized path coefficients of the estimated structural model are presented below. Figure 3 illustrates the relative magnitude of each structural relationship among the latent constructs.

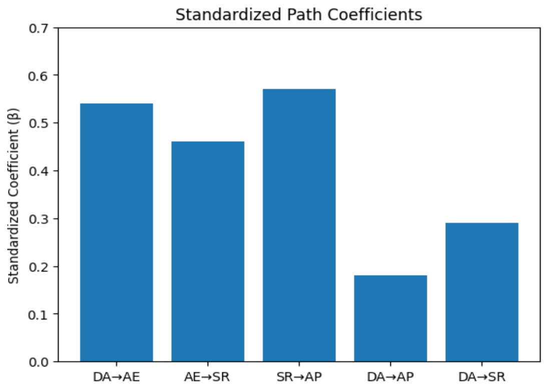

Figure 3: Standardized path coefficients.

3.2 Structural Path Coefficients

Figure 3 presents the standardized path coefficients of the structural model, illustrating the relative strength of the relationships among the latent variables. The most pronounced effect is observed for the relationship between self-regulation and academic performance (SR → AP; $\beta = 0.57$), confirming the central role of metacognitive and behavioral strategies in achieving high academic outcomes [16], [23].

A substantial effect is also observed for the relationship between digital activity and academic engagement (DA → AE; β = 0.54). This finding aligns with Learning Analytics frameworks suggesting that the intensity of interaction with LMS platforms contributes to higher levels of student participation in the learning process [3], [15].

Furthermore, the relationship between academic engagement and self-regulation (AE → SR; β = 0.46) supports the sequential logic of the proposed model, indicating that active engagement in learning activities facilitates the development of self-regulated learning behaviors.

The direct effect of digital activity on academic performance (DA → AP; β = 0.18) is statistically significant but considerably weaker than the indirect pathways mediated through engagement and self-regulation. The presence of the path DA → SR (β = 0.29) indicates that digital behavior partially contributes to the development of self-regulated learning skills. Nevertheless, the results demonstrate that the internal organization of learning processes remains the primary determinant of academic success.

The visualization of standardized coefficients enables a clear comparison of the relative strength of structural relationships and confirms that self-regulation is the strongest predictor of academic performance in the proposed SEM framework [11].

3.3 Model Fit Indices

Table 2 presents the key model fit indices of the structural equation model (SEM), which evaluate the extent to which the theoretically specified model corresponds to the empirical data.

The obtained χ^2/df value falls within the recommended range (< 3), indicating an acceptable balance between model complexity and explanatory power [12]. The incremental fit indices CFI and TLI exceed the threshold of 0.90, suggesting that the model provides a substantially better fit than the null model.

Table 2: SEM model fit indices.

Index	Value
χ^2/df	2.41
CFI	0.94
TLI	0.92
RMSEA	0.048
SRMR	0.041

The RMSEA value is below the recommended range of 0.06-0.08, indicating a high degree of approximation of the model to the population covariance structure. Additionally, the SRMR value is below 0.08, confirming a low level of residual discrepancies between observed and model-implied covariances [12].

The obtained indices meet widely accepted methodological thresholds in structural equation modeling literature [12]. The combination of multiple goodness-of-fit measures supports the statistical robustness and adequacy of the structural model.

Using several fit criteria is consistent with contemporary SEM evaluation guidelines and reduces the risk of biased interpretation based on a single metric [12]. Overall, the structural model demonstrates a good level of model fit, allowing for reliable interpretation of both direct and indirect effects and supporting the predictive potential of Learning Analytics in explaining academic performance.

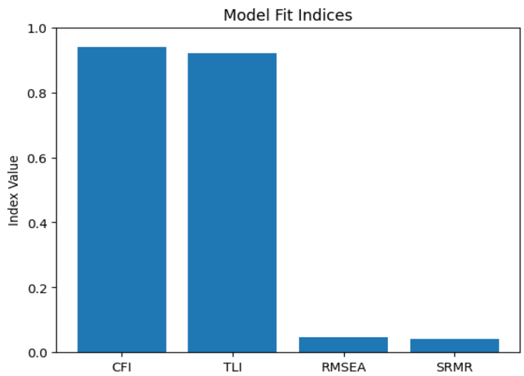

Figure 4: Model fit indices.

Figure 4 visualizes the key goodness-of-fit indices of the structural equation model (SEM), reflecting the degree of correspondence between the theoretically specified model and the empirical data. The values of the comparative fit indices CFI (0.94) and TLI (0.92) exceed the recommended threshold of 0.90, indicating a good model fit relative to the null model [12].

The low values of RMSEA (0.048) and SRMR (0.041) further confirm a minimal level of discrepancy between the observed covariance matrix and the model-implied covariance structure, satisfying the criteria for high model approximation [11], [20].

The combination of these indices demonstrates the statistical robustness of the structural configuration and confirms the correctness of the specified relationships among the latent variables. The visual representation allows a clear comparison of the obtained values with established threshold criteria and strengthens the interpretative justification of the results. Thus, the structural model can be considered adequate for analyzing both direct and indirect effects and for further interpreting the predictive role of Learning Analytics in explaining academic performance [11].

3.4 Explained Variance (R^2)

The proportion of explained variance for the endogenous constructs is as follows:

1) R^2 (Academic Engagement, AE) = 0.29.
2) R^2 (Self-Regulation, SR) = 0.41.
3) R^2 (Academic Performance, AP) = 0.52.

The model explains 52.0% of the variance in academic performance, which is considered a high explanatory power in educational research contexts [8].

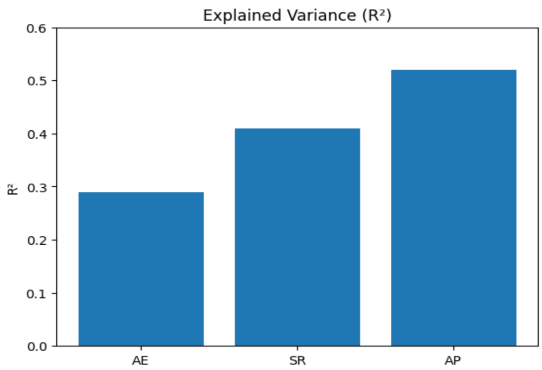

Figure 5: Explained variance (R^2).

Figure 5 presents the proportion of explained variance (R^2) for the endogenous latent variables in the structural model. The highest value is observed for academic performance (R^2 = 0.52), indicating that more than half of the variation in final academic outcomes is explained by the predictors included in the model-self-regulation and digital activity. This level of explained variance is considered high in educational analytics research and confirms the substantial explanatory power of the proposed model [8].

Self-regulation demonstrates a moderate level of explained variance (R^2 = 0.41), indicating a significant contribution of academic engagement and digital activity to the development of students' learning strategies.

The lowest R^2 value is observed for academic engagement (R^2 = 0.29). Nevertheless, this value remains statistically meaningful and corresponds to a moderate level of explanation in social science research [12]. This finding suggests that while digital activity is an important factor influencing engagement, it is not the sole determinant, which is consistent with the multidimensional nature of the engagement construct [14], [15].

The visual representation of R^2 values allows for a clear comparison of the relative contribution of predictors to each endogenous variable and confirms the structural logic of the model, where self-regulation functions as the central mechanism in the achievement of academic performance [16], [23].

3.5 Mediation Effect Analysis

Bootstrapping analysis with 5000 resamples was conducted to test the mediation hypothesis.

The indirect effect of the sequential pathway:

$$DA \rightarrow AE \rightarrow SR \rightarrow AP,$$

was estimated as:

$$\beta\_indirect = 0.14, 95\%\ CI\ [0.09; 0.21].$$

The confidence interval does not include zero, confirming the presence of a statistically significant partial mediation effect [12]. This result indicates that digital activity influences academic performance not only directly but also indirectly through the sequential mechanisms of academic engagement and self-regulation.

The main findings of the results section can be summarized as follows:

1) The measurement model demonstrated high reliability and validity.
2) The structural model exhibited good model fit.
3) Self-regulation emerged as the strongest predictor of academic performance.
4) Digital activity exerts both direct and indirect effects on academic outcomes.
5) The proposed model explains more than half of the variance in students' academic performance.

Figure 6 illustrates the magnitude of the indirect (mediated) effect of digital activity on academic performance, estimated using bootstrapping with5000 resamples and presented with a 95%

confidence interval. The estimated indirect effect is β = 0.14, with a confidence interval of [0.09; 0.21], which does not include zero. This indicates the statistical significance of the mediation mechanism [12].

These findings support the hypothesis that the influence of digital activity on academic outcomes is primarily transmitted through intermediate psychological variables-academic engagement and self-regulation.

The visualization of the confidence interval provides a clear representation of the robustness and positive direction of the effect. The absence of intersection with zero indicates the reliability of the mediating relationship and reduces the likelihood of random interpretation of the results [12].

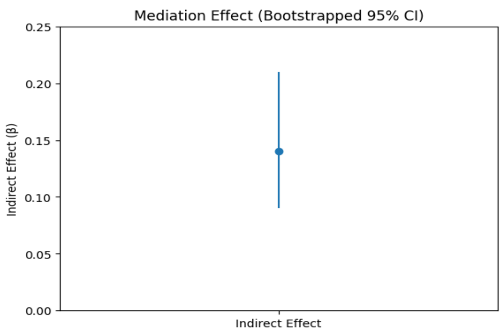

Figure 6: Mediation effect.

Thus, the figure confirms the presence of partial mediation: digital activity affects academic performance both directly and indirectly, while the mediated pathway through self-regulation plays a conceptually significant role within the overall structural model [16], [23].

4 DISCUSSION

4.1 Interpretation of Key Results

The findings confirm the theoretical validity of integrating Learning Analytics and Structural Equation Modeling for predicting academic performance. The structural model demonstrated strong model fit indices (CFI = 0.94; RMSEA = 0.048), indicating the adequacy of the proposed conceptual architecture [12].

Self-regulation emerged as the most significant predictor of academic performance (β = 0.57), which is consistent with extensive research in the field of self-regulated learning [16], [23]. This result suggests that the key mechanism transforming digital activity into academic achievement is not merely the intensity of LMS interaction but rather the student's ability to effectively organize and regulate their learning processes.

Digital activity demonstrated both direct and indirect effects on academic performance. The direct effect (β = 0.18) indicates that students who interact more actively with the LMS tend to achieve higher academic outcomes, consistent with findings from Learning Analytics research [3].

However, the indirect effect through academic engagement and self-regulation (β_indirect = 0.14) proved to be more substantial. This suggests that digital traces alone do not guarantee academic success; rather, their predictive value emerges only when combined with psychological mechanisms of learning regulation.

4.2 The Role of Academic Engagement as an Intermediate Mechanism

Academic engagement showed a significant effect on self-regulation (β = 0.46), supporting the hypothesis that active participation in the learning environment contributes to the development of metacognitive strategies [15], [24].

This finding allows engagement to be interpreted not only as a behavioral indicator but also as a transitional mechanism linking digital activity to stable learning strategies.

Within digital education contexts, engagement plays a crucial role: participation in forums, interaction with instructors and peers, and discussion of learning materials create opportunities for reflection and planning of learning activities [14], [15].

Thus, the model demonstrates a multilevel influence structure, where the behavioral layer (DA) transitions into the cognitive layer (SR) through the socio-behavioral mechanism of engagement (AE).

4.3 Comparison with International Studies

The obtained R^2 value for academic performance (0.52) is comparable to or exceeds the results reported in international studies, where explained variance typically ranges from 35% to 48% [8]. This indicates the strong explanatory capacity of the proposed model.

The findings are also consistent with the results of Broadbent and Poon [21], who reported that self-regulation is a stronger predictor of academic performance than simple measures of digital activity.

Our results confirm that metrics such as login frequency or time spent in the LMS are not sufficient predictors of academic success unless they are accompanied by strategic learning regulation.

Additionally, the observed partial mediation effect corresponds to theoretical models of learning regulation chains, where the digital environment functions as the context while psychological mechanisms act as the primary drivers of academic achievement [16], [17].

4.4 Theoretical Implications

This study expands existing Learning Analytics models by incorporating latent psychological constructs into structural analysis.

Unlike machine learning models that focus exclusively on predictive accuracy, SEM enables researchers to uncover causal mechanisms and clarify the theoretical nature of relationships [10].

Two major theoretical contributions can be highlighted:

1) Confirmation of a hierarchical influence structure: Digital Activity → Engagement → Self-Regulation → Academic Performance.
2) Demonstration that self-regulation functions as the central mediating mechanism, enhancing the interpretative value of Learning Analytics.

Thus, the proposed model integrates behavioral, social, and cognitive levels of analysis, aligning with contemporary interdisciplinary trends in educational research [3], [15].

4.5 Practical Implications

The findings have important practical implications for universities and developers of educational platforms.

Early Warning Systems. The model enables the identification of at-risk students not only through low digital activity but also through insufficient levels of self-regulation.

Personalized Learning. The mediation analysis highlights the importance of developing tools that support self-regulated learning, such as progress trackers, reminders, and adaptive recommendations.

LMS Design. Enhancing interactive elements within LMS platforms may increase engagement and indirectly improve academic performance.

Thus, integrating SEM into Learning Analytics provides a foundation for interpretable and scientifically grounded decision-making tools in educational policy.

4.6 Study Limitations

Despite the strong model fit, several limitations should be acknowledged:

1) the use of a simulated dataset;
2) the cross-sectional design of the study;
3) linear specification of relationships;
4) the absence of individual cognitive differences.

Future studies may extend the model using longitudinal data and nonlinear analytical approaches.

4.7 Future Research Directions

Promising directions for future research include:

1) integrating machine learning and SEM (hybrid modeling);
2) analyzing the temporal dynamics of digital activity;
3) incorporating motivational and affective variables;
4) conducting cross-cultural studies.

5 CONCLUSIONS

This study aimed to develop and empirically test a structural model for predicting academic performance based on Learning Analytics data using Structural Equation Modeling (SEM).

The results confirm the theoretical and statistical validity of the proposed model, integrating digital activity, academic engagement, self-regulation, and academic performance into a unified causal framework. The measurement model demonstrated strong reliability and validity indicators (α, CR, AVE), while the structural component exhibited good model fit according to key indices (CFI, TLI, RMSEA, SRMR), meeting contemporary standards for research in educational analytics and SEM [12].

A key finding of the study is the confirmation of the central role of self-regulation as the strongest predictor of academic performance. Although digital activity remains important, its influence on academic outcomes is primarily mediated through psychological mechanisms-engagement and self-regulated learning.

Thus, students' digital traces within LMS environments gain predictive value only when interpreted in the context of deeper cognitive and behavioral processes [16], [23].

The identified partial mediation effect highlights the multilevel nature of educational dynamics and emphasizes the importance of interpreting Learning Analytics not only as a monitoring tool but also as a means of understanding learning behavior mechanisms [3], [15].

From a practical perspective, the proposed model can support the development of early warning systems for academic risk detection and personalized learning support. Unlike "black box" machine learning algorithms, structural modeling provides interpretability and transparency, which are particularly important in educational contexts [10], [28].

Universities may use such models to identify students with low levels of self-regulation and design targeted interventions aimed at fostering effective learning strategies.

Despite its contributions, the study has certain limitations. The use of simulated data limits the direct generalizability of findings to specific institutional contexts. Additionally, the cross-sectional design prevents definitive causal inference. Future research incorporating longitudinal datasets could provide deeper insights into the dynamics of digital activity and its transformation into academic outcomes.

Overall, the results demonstrate that the integration of Learning Analytics and structural modeling represents a theoretically grounded and promising approach to predicting academic performance. The proposed model not only explains more than half of the variance in academic outcomes but also reveals the underlying mechanisms through which digital behaviors are transformed into educational achievements.

This contributes to the advancement of educational analytics and the development of evidence-based decision-support systems in higher education.

REFERENCE

[1] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," EDUCAUSE Review, vol. 46, no. 5, pp. 30-40, 2011.

[2] R. Ferguson, "Learning analytics: Drivers, developments and challenges," International Journal of Technology Enhanced Learning, vol. 4, no. 5-6, pp. 304-317, 2012, [Online]. Available: https://doi.org/10.1504/IJTEL.2012.051816.

[3] G. Siemens, "Learning analytics: The emergence of a discipline," American Behavioral Scientist, vol. 57, no. 10, pp. 1380-1400, 2013, [Online]. Available: https://doi.org/10.1177/0002764213498851.

[4] K. E. Arnold and M. D. Pistilli, "Course signals at Purdue: Using learning analytics to increase student success," in Proceedings of LAK, 2012, pp. 267-270.

[5] D. T. Tempelaar, B. Rienties, and B. Giesbers, "In search for the most informative data for feedback generation: Learning analytics in a data-rich context," Computers in Human Behavior, vol. 47, pp. 157-167, 2015, [Online]. Available: https://doi.org/10.1016/j.chb.2014.05.038.

[6] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Transactions on Systems, Man, and Cybernetics, vol. 40, no. 6, pp. 601-618, 2010.

[7] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, Multivariate Data Analysis, 7th ed., Pearson, 2014.

[8] D. T. Tempelaar, B. Rienties, and Q. Nguyen, "Learning analytics in higher education," Computers in Human Behavior, vol. 69, pp. 375-386, 2017.

[9] D. Gašević, S. Dawson, and G. Siemens, "Let's not forget: Learning analytics are about learning," TechTrends, vol. 59, no. 1, pp. 64-71, 2015.

[10] S. Slade and P. Prinsloo, "Learning analytics: Ethical issues and dilemmas," American Behavioral Scientist, vol. 57, no. 10, pp. 1510-1529, 2013.

[11] R. B. Kline, Principles and Practice of Structural Equation Modeling, 4th ed., Guilford Press, 2016.

[12] L. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis," Structural Equation Modeling, vol. 6, no. 1, pp. 1-55, 1999.

[13] D. Ifenthaler and J. Y. K. Yau, "Utilising learning analytics for study success: Reflections on current empirical findings," British Journal of Educational Technology, vol. 51, no. 5, pp. 1744-1760, 2020.

[14] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept," Review of Educational Research, vol. 74, no. 1, pp. 59-109, 2004.

[15] C. R. Henrie, L. R. Halverson, and C. R. Graham, "Measuring student engagement in technology-mediated learning," Computers & Education, vol. 90, pp. 36-53, 2015.

[16] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," Theory Into Practice, vol. 41, no. 2, pp. 64-70, 2002.

[17] J. Broadbent and W. L. Poon, "Self-regulated learning strategies & academic achievement in online higher education," Internet and Higher Education, vol. 27, pp. 1-13, 2015.

[18] M. Richardson, C. Abraham, and R. Bond, "Psychological correlates of university students' academic performance," Psychological Bulletin, vol. 138, no. 2, pp. 353-387, 2012.

[19] A. Bandura, Social Foundations of Thought and Action, Prentice-Hall, 1986.

[20] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," in Handbook of Self-Regulation, 1998, pp. 277-304.

[21] J. Broadbent, “Comparing online and blended learner’s self-regulated learning strategies,” Internet and Higher Education, vol. 33, pp. 24-32, 2017.

[22] W. Xing and D. Du, “Dropout prediction in MOOCs,” in Proceedings of LAK, 2019, pp. 60-65.

[23] E. Panadero, “A review of self-regulated learning,” Educational Research Review, vol. 22, pp. 1-15, 2017.

[24] P. R. Pintrich, “The role of motivation in promoting and sustaining self-regulated learning,” International Journal of Educational Research, vol. 31, no. 6, pp. 459-470, 1999.

[25] G. D. Kuh, “The National Survey of Student Engagement,” Change, vol. 33, no. 3, pp. 24-32, 2001.

[26] M. Credé and N. R. Kuncel, “Study habits, skills, and attitudes,” Perspectives on Psychological Science, vol. 3, no. 6, pp. 425-453, 2008.

[27] D. Gašević, N. Mirriahi, and P. Long, “Analytics of the future,” Journal of Learning Analytics, vol. 1, no. 2, pp. 1-4, 2014.

[28] R. S. J. d. Baker and P. S. Inventado, “Educational data mining and learning analytics,” in Learning Analytics, Springer, 2014, pp. 61-75.