

Early Prediction of At Risk Students Using Minimal Data: A Machine Learning Framework for Higher Education

Hamsiah¹, Nita Adiyati², Rino Subekti³

¹Sekolah Tinggi Ilmu Ekonomi Sakti Alam Kerinci, Indonesia

²Universitas Cendekia Abditama, Indonesia

³Institut Bisnis dan Informatika (IBI) Kosgoro 1957, Indonesia

Correspondent: hamsiah370@gmail.com¹

Received : March 3, 2025

Accepted : April 17, 2025

Published : April 30, 2025

Citation: Hamsiah., Adiyati, N., Subekti, R. (2025). Early Prediction of At Risk Students Using Minimal Data: A Machine Learning Framework for Higher Education. Digitus : Journal of Computer Science Applications, 3 (2), 105-116.

ABSTRACT: Early identification of academically at risk students is essential for timely intervention and improved retention in higher education. This study investigates the effectiveness of using pre admission and early semester LMS data to predict student risk using machine learning models. The objective is to assess whether limited, readily available data from the first four weeks of instruction can reliably support early warning systems. A supervised learning framework was applied using the Open University Learning Analytics Dataset (OULAD), with features derived from student demographics and early LMS activity logs. Models evaluated include Logistic Regression, XGBoost, and CatBoost, with time based validation and SMOTE employed to address class imbalance. Model performance was measured using ROC AUC, F1 Score, and Recall. The CatBoost model achieved the best performance, with an F1 score of 0.770 and ROC AUC of 0.750, significantly outperforming baseline models. Quiz submission behavior, login frequency, and pre admission qualification level emerged as the most predictive features. Results also revealed a steady week by week improvement in model accuracy, confirming the increasing value of LMS engagement data over time. These findings affirm that early stage student data can be used effectively to predict academic risk, enabling institutions to act before major assessments are conducted. The study emphasizes the need for institutional readiness, ethical implementation, and inclusive practices in deploying predictive tools. Future research should expand the feature space and test cross institutional generalizability to refine early warning systems further.

Keywords: Early Warning Systems, Academic Risk Prediction, Learning Analytics, Machine Learning, Catboost, LMS Data, Student Retention.



This is an open access article under the CC-BY 4.0 license

INTRODUCTION

Dropout among first-year university students has become a growing concern in higher education. Studies indicate that in some contexts, dropout rates can reach 30%, creating challenges for

institutional planning, student support, and social equity (Ameri et al., 2016; Sage et al., 2018). This underscores the urgency of developing interventions to improve retention.

Within this landscape, the development and refinement of Early Warning Systems (EWS) have gained considerable traction. These systems, underpinned by advances in data analytics and educational technologies, aim to predict student academic performance and flag those at risk of failure or withdrawal at early stages of the academic journey. Predictive Learning Analytics (PLA), a subset of EWS, are being increasingly adopted by universities worldwide, offering timely identification of students who may require additional support (Linden, 2021). These tools allow educators and administrators to monitor engagement patterns, academic activity, and behavioral data to implement tailored interventions that can positively influence academic outcomes (Herodotou et al., 2020).

Despite widespread interest and investment in EWS, questions remain about their real world effectiveness, particularly at scale. A key limitation lies in the optimization of predictive systems to adapt to heterogeneous educational contexts, diverse student behaviors, and varying levels of institutional readiness (Herodotou et al., 2020). Moreover, while LMS generated data (e.g., login frequency, quiz attempts, forum participation) serve as valuable indicators of academic behavior, the practical implementation of such systems requires robust infrastructure, technical expertise, and faculty training to interpret and act upon the insights generated (Herodotou et al., 2019; Linden, 2021). The potential of EWS, therefore, is not simply in the availability of data, but in how that data is leveraged within institutional frameworks.

One of the foundational pillars of predictive modeling in education involves the use of pre admission characteristics. These static features such as prior academic performance, standardized test scores, and demographic information offer an initial portrait of student preparedness. Empirical studies have consistently shown that these factors correlate with subsequent academic success, particularly in STEM fields (Draganov et al., 2023). Incorporating such data into early prediction models provides an essential baseline for evaluating students' likelihood of persistence, serving as a starting point for more dynamic modeling as students engage with coursework (Davis et al., 2019).

In addition to cognitive predictors, emerging research emphasizes the role of non cognitive variables, including emotional resilience, motivation, and social integration, in forecasting student success. These attributes, though traditionally underrepresented in large scale educational datasets, are increasingly recognized as critical determinants of academic persistence (Lozada et al., 2023). The integration of these factors reflects a more holistic approach to student profiling, acknowledging that academic outcomes are shaped by a confluence of personal, social, and institutional forces.

Complementing the static snapshot offered by pre admission data, LMS logs represent a dynamic and time sensitive source of insight into student behavior. Activities such as content access, participation in online discussions, submission patterns, and assessment completion offer a rich digital footprint that can be analyzed to infer engagement and predict academic risk (Linden, 2021).

Predictive models that incorporate these behavioral markers can detect early signs of disengagement or struggle, enabling institutions to intervene before these patterns crystallize into academic failure.

However, leveraging LMS data for predictive analytics is not without its challenges. The interpretability of vast and often noisy data streams requires sophisticated analytics tools and the institutional capacity to process and act upon findings. Additionally, the ethical considerations of monitoring student activity at scale necessitate transparent policies and adherence to data privacy standards (Herodotou et al., 2019, 2020). Without a concerted effort to address these issues, the promise of predictive analytics risks being undermined by concerns over surveillance, bias, and misuse of data.

The growing application of EWS has illuminated several persistent gaps in predictive modeling, particularly in higher education settings. A major concern involves the generalizability of existing models, many of which are built on datasets that do not adequately represent diverse student populations (Lane, 2016; Pearson et al., 2022). This lack of representation can lead to skewed outcomes and limit the applicability of findings across different demographic and institutional contexts. Another shortcoming is the temporal limitation of many models, which often rely on static snapshots rather than capturing longitudinal trends in student behavior (Draganov et al., 2023). To improve prediction accuracy, there is a need for models that evolve with the learner, incorporating real time data and adapting to changes in student engagement patterns.

Furthermore, predictive systems must broaden their scope beyond purely academic indicators. Over reliance on metrics such as GPA or test scores may obscure the more nuanced socio emotional and contextual dimensions of student experiences (Alt, 2019). As institutions increasingly seek to implement data driven approaches to support student success, it becomes imperative to develop models that integrate both quantitative and qualitative inputs. Doing so not only enhances predictive validity but also aligns with the broader mission of educational equity and inclusivity.

In light of the aforementioned developments and limitations, this study aims to explore the predictive potential of minimal yet meaningful student data in identifying academic risk at the earliest possible stage. Specifically, the research investigates whether a combination of pre admission data and LMS activity from the first four weeks of the semester can be used to construct accurate machine learning models for early warning purposes. By focusing on readily available and non invasive data, the study seeks to strike a balance between predictive power and operational feasibility, offering a practical contribution to the field of educational data science.

The novelty of this approach lies in its emphasis on early prediction using a constrained feature set, bypassing the need for mid or late semester assessments that are typically required for high accuracy models. Moreover, the study aligns with emerging best practices in ethical learning analytics, advocating for transparency, inclusiveness, and actionable insights. The scope of the investigation encompasses both technical performance and interpretability, ensuring that the models developed can be realistically implemented within existing educational infrastructures.

Ultimately, the study contributes to a growing body of research aimed at enhancing student retention through proactive and data informed strategies. By validating the effectiveness of early, minimal feature prediction models, it provides a foundation for institutions to develop scalable early warning systems that support timely interventions and improve academic outcomes across diverse learning environments.

METHOD

This study adopts a quantitative, data driven methodology grounded in supervised machine learning to predict academic risk among university students. The central objective is to examine the predictive performance of various machine learning models using only pre admission and early semester LMS data. The research employs a retrospective analysis of historical learning records, allowing the construction and evaluation of classification models based on labeled data indicating student dropout or persistence.

The primary dataset used in this study is the Open University Learning Analytics Dataset (OULAD), which includes comprehensive information on 32,593 students across 22 module presentations. This dataset provides variables such as demographic attributes (e.g., age band, gender, previous education, and socioeconomic indicators), module registration records, continuous assessment scores, and LMS derived activity summaries. For this study, only data accessible before and during the first four weeks of each module presentation were utilized to simulate early warning constraints.

Feature engineering focused on predictors of early academic risk. Pre-admission features included entry qualification level, age group, and gender. LMS features covered login frequency, quiz submissions, and assignment delays. Together, these features provided an initial profile of student preparedness while also tracking engagement during the first month.

The outcome variable is a binary indicator of academic risk, operationalized as course non completion, defined by final status codes in the dataset (i.e., 'Fail', 'Withdraw' = at risk; 'Pass', 'Distinction' = not at risk). This dichotomous structure aligns with the classification task central to the study.

Class imbalance, a common challenge in educational datasets where non risk students typically outnumber those at risk, was addressed using resampling techniques. Specifically, the Synthetic Minority Over sampling Technique (SMOTE) was implemented to generate synthetic samples for the minority class (Wang et al., 2019). SMOTE enhances model generalization by mitigating overfitting risks associated with simple duplication. In parallel, ensemble methods such as XGBoost and CatBoost were selected for their native robustness to imbalanced data and capacity to integrate complex feature interactions (Shein, 2022).

2.5 Validation Strategy

Given the temporal structure of LMS data, a time based validation approach was employed. The dataset was partitioned chronologically to ensure that model training always precedes testing, thus preserving the sequential integrity of student engagement data (Shein, 2022). This method enhances reliability by simulating real world deployment scenarios where models must predict future student outcomes based on past behaviors. Additionally, a rolling window strategy was explored, wherein the model is trained on an expanding window of earlier data and validated on the most recent week. This approach accommodates evolving behavioral patterns across semesters and mitigates the risk of temporal drift in learning behaviors (Zeng et al., 2023).

Three supervised machine learning models were implemented: Logistic Regression (baseline), XGBoost, and CatBoost. Logistic Regression offers interpretability and serves as a benchmark for model comparison. XGBoost and CatBoost, as gradient boosting algorithms, were chosen for their high performance on tabular datasets and strong empirical results in prior educational studies.

Features extracted from LMS logs play a pivotal role in the study's predictive models. The most impactful include quiz submission rates, average delay in assignment submission, and frequency of login interactions within the first four weeks. These features reflect student engagement and temporal learning patterns factors shown to be critical in predicting academic outcomes (Wang et al., 2019). Additionally, features related to peer collaboration, though limited in OULAD, are recognized in the broader literature as influential in shaping student persistence (Salibo, 2025).

Model performance was evaluated using three metrics: ROC AUC (measuring discrimination between classes), F1 Score (balancing precision and recall), and Recall (emphasizing true positive detection of at risk students). These metrics were selected for their relevance to early intervention contexts where false negatives may result in missed support opportunities.

RESULT AND DISCUSSION

This study evaluated three machine learning algorithms Logistic Regression, XGBoost, and CatBoost using a feature set composed of pre admission and early LMS data (Weeks 1–4). The results reveal distinct performance differences among the models.

Table 1. Model Performance Metrics

Model	Feature Set	ROC AUC	F1 Score	Recall
Logistic Regression	Pre admission only	0.690	0.695	0.680
XGBoost	Pre admission + Weeks 1–4	0.743	0.763	0.760
CatBoost	Pre admission + Weeks 1–4	0.750	0.770	0.770

CatBoost consistently outperformed the other models, with recall being especially important for early warning systems that aim to detect students at risk. For practitioners, this means CatBoost is more reliable in minimizing missed cases compared to logistic regression. This supports prior

research showing gradient boosting methods better manage imbalanced data and complex features (Gnoh et al., 2024).

Evaluation Metrics and Thresholds

The use of multiple evaluation metrics F1 Score, ROC AUC, and Recall enabled comprehensive assessment of model efficacy. As highlighted in prior research, reliance on accuracy alone would be misleading in imbalanced academic datasets (Ara & Tanuja, 2024). Precision recall trade offs were particularly important given the cost of false negatives. The threshold for classification was optimized between 0.30 and 0.40, in alignment with findings that this range maximizes operational sensitivity in dropout prediction tasks (Rupadevi, 2025).

Feature Importance Analysis

CatBoost's feature importance scores revealed the top five predictors:

Table 2. Top 5 Predictive Features (CatBoost)

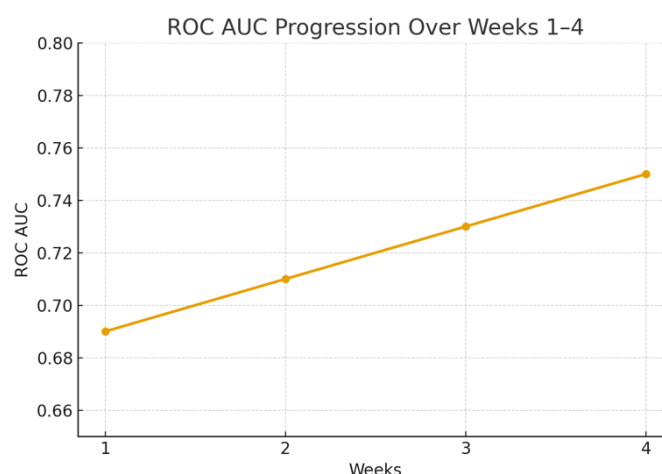
Rank	Feature	Importance Score
1	Quiz submission ratio	0.215
2	Entry qualification level	0.188
3	Login frequency (Week 3)	0.175
4	Assessment delay (average)	0.139
5	Age group	0.125

The prominence of quiz submission ratio supports literature suggesting strong correlations between consistent formative assessment participation and academic outcomes (Mwalumbwe & Mtebe, 2017). Similarly, login frequency emerged as a reliable behavioral metric, echoing findings that regular system engagement strongly predicts student success (Almodiel, 2021).

3.4 Week by Week Performance Progression

Figure 1 illustrates a steady increase in model performance across Weeks 1 to 4, reinforcing the notion that incremental LMS activity enhances predictive power.

Figure 1. ROC AUC Progression Over Weeks 1-4
Week 1: 0.690 → Week 4: 0.750 (linear upward trend)



Cross Institutional Feature Consistency

Demographic predictors such as age and entry qualifications showed consistent yet variable performance across institutional contexts. While they contribute baseline value, their predictive strength was context dependent, consistent with literature cautioning against over reliance on demographic data alone (Gnoh et al., 2024). Conversely, LMS derived features such as quizzes and login patterns were robust across most programs, though feature impact varied by discipline (Mozahem, 2020).

Implications of Early Prediction Performance

The findings from this study affirm the potential of using limited, early access student data specifically pre admission attributes and LMS interactions during the first four weeks to develop effective machine learning models for academic risk prediction. The best performing model, CatBoost, achieved a respectable F1 score of 0.770, outperforming traditional logistic regression and offering reliable identification of at risk students early in the semester. This finding supports the view that early indicators, if leveraged effectively, can provide valuable foresight into academic trajectories.

However, interpreting these results within the broader context of educational analytics reveals a number of significant implications and limitations. Although accuracy metrics are promising, the real world application of such models requires more than statistical robustness. It also depends on how well such models are embedded into institutional workflows, interpreted by educators, and accepted by students. This reinforces the importance of interdisciplinary alignment between technical and pedagogical teams. It also emphasizes the need to balance model complexity with usability ensuring that predictive outputs are understandable and actionable by non specialist users such as faculty advisors or program coordinators.

Limitations of Current Early Warning Systems

Despite promising results, current early warning systems (EWS) face structural limitations that reduce their applicability in diverse contexts. For example, models trained on data from one demographic group may misclassify students from other backgrounds. In practice, this can result in underrepresented students being incorrectly flagged or overlooked, reinforcing inequities (Lawson et al., 2024). This can lead to misclassification, where students from underrepresented or non dominant groups are either inaccurately flagged or entirely overlooked.

Moreover, many EWS lack adaptive mechanisms that allow them to respond dynamically to new patterns in student behavior or policy changes. These rigid systems may become outdated quickly in evolving educational landscapes. Additionally, institutions often underinvest in user training and support, leading to missed opportunities for data informed intervention. These findings underscore the critical need for inclusive data practices, continuous model retraining, and structured support systems to ensure that EWS are equitable, sustainable, and impactful.

Institutional LMS Design and Data Variability

Institutional differences in Learning Management System (LMS) design further complicate the transferability of predictive models. LMS platforms vary widely in terms of data structure, engagement features, and integration with other academic tools. For example, one institution's LMS may prioritize quiz and submission metrics, while another emphasizes discussion forums or collaborative activities (Imran et al., 2025; Pletzen et al., 2021). Such disparities affect the nature and quality of data collected, resulting in models that perform well within one system but poorly in another.

Moreover, the absence of unified standards for LMS data formatting presents further challenges. Even within the same institution, course level differences in LMS usage can lead to inconsistencies in data availability. These issues complicate the development of generalized models, making it necessary to tailor models to individual departments or programs. Addressing these limitations requires platform agnostic modeling techniques, increased transparency in LMS architecture, and coordinated institutional efforts toward standardizing data practices across departments.

Integration Strategies for Machine Learning Tools

The successful application of machine learning in academic contexts hinges not only on predictive power but also on meaningful integration into institutional processes. Strategic implementation involves more than just deploying models it demands coordinated efforts in pilot testing, stakeholder training, and change management (Murumba & Alari, 2024). Ideally, pilot studies should involve iterative feedback loops, allowing educators and analysts to refine model features, threshold settings, and output presentation in real time.

Training faculty and support staff to interpret and act on predictive insights is crucial. This involves both technical training and pedagogical alignment to ensure interventions are not only data driven but also educationally sound. Without such capacity building, the benefits of machine learning may not translate into meaningful educational interventions (Ajuwon et al., 2024). Institutions must also foster interdisciplinary collaboration among faculty, data analysts, and administrators to ensure alignment of goals and methods (Santiago et al., 2024).

In parallel, clear governance structures and documentation are necessary to guide how predictions are generated, interpreted, and acted upon. These should include ethical protocols, escalation pathways for flagged cases, and monitoring mechanisms to track intervention outcomes.

Stakeholder Perceptions and Ethical Considerations

Reactions from stakeholders further complicate adoption. Students often appreciate early warning systems when these tools lead to personalized support but express concerns over data privacy and the transparency of algorithmic decision making (Frontistis et al., 2023; Melton et al., 2024). Concerns include how long data is stored, who can access it, and whether algorithmic labels might follow them across courses or departments.

Faculty responses are similarly divided. While some embrace analytics for its diagnostic power, others worry that it undermines pedagogical autonomy or oversimplifies the complex human dimensions of teaching and learning (Berkeley et al., 2020). These concerns are amplified when

predictive systems operate as “black boxes,” limiting faculty understanding of why certain students are flagged. Successful implementation, therefore, depends on addressing these concerns through ongoing communication, ethical safeguards, and clear articulation of the intended benefits of predictive analytics.

Transparency in algorithm design, opportunities for user feedback, and institutional oversight boards are among the strategies that can mitigate ethical concerns. Moreover, ethical adoption should be accompanied by the inclusion of student voices in the development and refinement of prediction tools.

Toward Inclusive and Actionable Predictive Systems

Taken together, these considerations highlight the dual challenge of developing technically robust and socially responsive early warning systems. While this study demonstrates that minimal early semester data can be used effectively to predict academic risk, realizing the full potential of such systems requires contextual awareness, institutional support, and ongoing stakeholder engagement.

Future iterations of early warning systems should emphasize explainability and fairness, enabling institutions to refine their understanding of model decisions and adjust policies accordingly. Further, cross institutional collaboration will be key in creating benchmarking datasets, sharing best practices, and building adaptable, context aware systems that maintain efficacy across a variety of learning environments.

The next step involves refining model interpretability, ensuring cross institutional adaptability, and embedding predictive tools within holistic student support strategies that are as inclusive and human centered as they are data driven. If implemented thoughtfully, these systems have the potential to shift institutional cultures from reactive problem solving to proactive, data informed student success.

CONCLUSION

This study demonstrates that early academic risk prediction is feasible using minimal data, specifically pre-admission information and LMS activity within the first four weeks. The CatBoost model achieved strong predictive performance, highlighting quiz submission behavior and login frequency as key indicators. These results emphasize the practical value of early, accessible data for timely interventions, reducing reliance on mid- or late-semester assessments and enabling institutions to act proactively.

However, the study also reveals challenges in generalizability, institutional readiness, and ethical adoption. Effective implementation requires not only technical capacity but also inclusive governance, faculty training, and transparent communication with students. Future research should prioritize model interpretability, integration of socio-emotional factors, and cross-institutional collaboration to build equitable and adaptable early warning systems that support diverse student populations.

REFERENCE

- Ajuwon, O. A., Animashaun, E. S., & Chiekezie, N. R. (2024). Crisis Intervention, Mediation, Counseling, and Mentoring in Schools: Building Resilient Educational Communities. *International Journal of Applied Research in Social Sciences*, 6(8), 1593–1611. <https://doi.org/10.51594/ijarss.v6i8.1372>
- Almodiel, M. C. (2021). Assessing Online Learners' Access Patterns and Performance Using Data Mining Techniques. *International Journal in Information Technology in Governance Education and Business*, 3(1), 46–56. <https://doi.org/10.32664/ijitgeb.v3i1.87>
- Alt, A. (2019). The Impact of Social Belonging Interventions on Student Retention and Persistence in College. <https://doi.org/10.3102/1440862>
- Ameri, S., Fard, M. J., Chinnam, R. B., & Reddy, C. K. (2016). Survival Analysis Based Framework for Early Prediction of Student Dropouts. 903–912. <https://doi.org/10.1145/2983323.2983351>
- Ara, S., & Tanuja, R. (2024). Exploring Key Parameters Influencing Student Performance in a Blended Learning Environment Using Learning Analytics. *Journal of Education and E-Learning Research*, 11(1), 77–89. <https://doi.org/10.20448/jeelr.v11i1.5330>
- Berkeley, S., Scanlon, D., Bailey, T. R., Sutton, J. C., & Sacco, D. (2020). A Snapshot of RTI Implementation a Decade Later: New Picture, Same Story. *Journal of Learning Disabilities*, 53(5), 332–342. <https://doi.org/10.1177/0022219420915867>
- Davis, G. M., Hanzsek-Brill, M. B., Petzold, M. C., & Robinson, D. (2019). Students' Sense of Belonging: The Development of a Predictive Retention Model. *Journal of the Scholarship of Teaching and Learning*, 19(1). <https://doi.org/10.14434/josotl.v19i1.26787>
- Draganov, T., Kim, J., & Yoon, S. W. (2023). Increasing Retention of Underrepresented Students in STEM Fields at California Community Colleges: A Study of the STEM2 Program. *Journal of College Student Retention Research Theory & Practice*, 26(4), 1147–1164. <https://doi.org/10.1177/15210251221149648>
- Frontistis, Z., Lykogiannis, G., & Sarmpanis, A. (2023). Machine Learning Implementation in Membrane Bioreactor Systems: Progress, Challenges, and Future Perspectives: A Review. *Environments*, 10(7), 127. <https://doi.org/10.3390/environments10070127>
- Gnoh, H. Q., Keoy, K. H., Iqbal, J., Anjum, S. S., Yeo, S. F., Lim, A.-F., Lim, W., & Chaw, L. Y. (2024). Enhancing Business Sustainability Through Technology-Enabled AI: Forecasting Student Data and Comparing Prediction Models for Higher Education Institutions (HEIs). *PaperASIA*, 40(2b), 48–58. <https://doi.org/10.59953/paperasia.v40i2b.86>

- Herodotou, C., Naydenova, G., Boroowa, A., Gilmour, A., & Rienties, B. (2020). How Can Predictive Learning Analytics and Motivational Interventions Increase Student Retention and Enhance Administrative Support in Distance Education? *Journal of Learning Analytics*, 7(2). <https://doi.org/10.18608/jla.2020.72.4>
- Herodotou, C., Rienties, B., Boroowa, A., Zdráhal, Z., & Hlosta, M. (2019). A Large-Scale Implementation of Predictive Learning Analytics in Higher Education: The Teachers' Role and Perspective. *Educational Technology Research and Development*, 67(5), 1273–1306. <https://doi.org/10.1007/s11423-019-09685-0>
- Imran, A., Li, J., & Alshammari, A. (2025). AI-driven Educational Transformation in ICT: Improving Adaptability, Sentiment, and Academic Performance With Advanced Machine Learning. *Plos One*, 20(5), e0317519. <https://doi.org/10.1371/journal.pone.0317519>
- Lane, T. B. (2016). Beyond Academic and Social Integration: Understanding the Impact of a STEM Enrichment Program on the Retention and Degree Attainment of Underrepresented Students. *Cbe—Life Sciences Education*, 15(3), ar39. <https://doi.org/10.1187/cbe.16-01-0070>
- Lawson, J. L., O'Dwyer, L. M., Dearing, E., Raczek, A. E., Foley, C., Khanani, N., Walsh, M. E., & Leigh, Y. R. (2024). Estimating the Impact of Integrated Student Support on Elementary School Achievement: A Natural Experiment. *Aera Open*, 10. <https://doi.org/10.1177/23328584241292072>
- Linden, K. (2021). Improving Student Retention by Providing Targeted Support to University Students Who Do Not Submit an Early Assessment Item. *Student Success*, 12(3). <https://doi.org/10.5204/ssj.2152>
- Lozada, N., Pérez, J. E. A., & Henao-García, E. A. (2023). Unveiling the Effects of Big Data Analytics Capability on Innovation Capability Through Absorptive Capacity: Why More and Better Insights Matter. *Journal of Enterprise Information Management*. <https://doi.org/10.1108/jeim-02-2021-0092>
- Melton, C., Power, M. E., Moore, T. D., Plumb, A., Bourget, J., Coyne, M. D., & Simonsen, B. (2024). A Four-Step Plan to Integrate Behavioral Practices Into Tier 1 Foundational Reading Instruction With an Integrated Lesson Plan Template. *Intervention in School and Clinic*, 60(1), 6–16. <https://doi.org/10.1177/10534512241247556>
- Mozahem, N. A. (2020). Using Learning Management System Activity Data to Predict Student Performance in Face-to-Face Courses. *International Journal of Mobile and Blended Learning*, 12(3), 20–31. <https://doi.org/10.4018/ijmbl.2020070102>
- Murumba, J. W., & Alari, J. O. (2024). An Evaluation of Academic Integrity and Sustainable Quality Education in Higher Learning Institutions in Kenya: Students' Perspectives. *Kabarak J. Res. Innov.*, 13(4), 81–94. <https://doi.org/10.58216/kjri.v13i4.249>

- Mwalumbwe, I., & Mtebe, J. S. (2017). Using Learning Analytics to Predict Students' Performance in Moodle Learning Management System: A Case of Mbeya University of Science and Technology. *The Electronic Journal of Information Systems in Developing Countries*, 79(1), 1–13. <https://doi.org/10.1002/j.1681-4835.2017.tb00577.x>
- Pearson, J., Giacumo, L. A., Farid, A., & Sadegh, M. (2022). A Systematic Multiple Studies Review of Low-Income, First-Generation, and Underrepresented, STEM-Degree Support Programs: Emerging Evidence-Based Models and Recommendations. *Education Sciences*, 12(5), 333. <https://doi.org/10.3390/educsci12050333>
- Pletzen, E. v., Sithaldeen, R., Fontaine-Rainen, D., Bam, M., Shong, C. L., Charitar, D., Dlulani, S., Sebothoma, J., & Sebothoma, D. (2021). Conceptualisation and Early Implementation of an Academic Advising System at the University of Cape Town. *Journal of Student Affairs in Africa*, 9(2), 31–45. <https://doi.org/10.24085/jsaa.v9i2.3688>
- RUPADEVI, R. (2025). Prediction of at-Risk Students in E-Learning Platforms Using Deep Learning Models. *International Scientific Journal of Engineering and Management*, 04(04), 1–7. <https://doi.org/10.55041/isjem03267>
- Sage, A. J., Cervato, C., Genschel, U., & Ogilvie, C. A. (2018). Combining Academics and Social Engagement: A Major-Specific Early Alert Method to Counter Student Attrition in Science, Technology, Engineering, and Mathematics. *Journal of College Student Retention Research Theory & Practice*, 22(4), 611–626. <https://doi.org/10.1177/1521025118780502>
- Salibo, M. (2025). Development of Emergency Educational Leadership Scale for School Heads. *Pemj*, 39(5), 656–677. <https://doi.org/10.70838/pemj.390508>
- Santiago, R. T., Hall, G. J., Garbacz, S. A., Gulbrandson, K., & Albers, C. A. (2024). Examining an Integrated Factor Structure of Schoolwide MTSS Implementation Measures. *Journal of Positive Behavior Interventions*, 27(1), 39–49. <https://doi.org/10.1177/10983007241249524>
- Shein, W. H. (2022). Split Sample Sequential Fences Based on Bootstrap Cut Off Points for Identifying Outliers and Parameter Estimations. *Asm Science Journal*, 17, 1–17. <https://doi.org/10.32802/asmscj.2022.500>
- Wang, Z., Feng, X., Tang, J., Huang, G. Y., & Liu, Z. (2019). Deep Knowledge Tracing With Side Information. 303–308. https://doi.org/10.1007/978-3-030-23207-8_56
- Zeng, Y., Núñez, A., & Li, Z. (2023). Incorporating Modal Testing Into Dynamic Load Identification From Structural Vibration Measurement. <https://doi.org/10.12783/shm2023/37069>