



A Literature Review of Empirical Studies on Predictive Learning Analytics for Student Success in Higher Education

Janry B. Gabuyo

Open University System | Pangasinan State University
Pangasinan, Philippines

Article Info:

Received: 01 June 2025; Revised: 09 Aug 2025; Accepted: 21 Oct 2025; Available Online: 15 Dec 2025

Abstract – Predictive Learning Analytics (PLA) has emerged as a practical approach for using data from digital learning environments to better understand student progress and to support timely academic assistance. This research reviews recent work on PLA in higher education with a focus on how prediction is commonly framed and what factors repeatedly limit dependable use. The synthesis shows that PLA studies typically build models from routine academic and engagement-related information, aiming to flag students who may require additional support. At the same time, the literature raises consistent concerns about whether commonly used indicators capture meaningful learning process, whether model remains dependable when moved to new courses or cohorts, and whether results are presented in a way that can be responsibly interpreted and acted upon. Evidence on practical value of early-warning system is uneven, particularly when predictors are not embedded in clear support procedures. Overall, it argues that PLA is most defensible when models are validated within the context, communicated transparently, and implemented as part of student-centered support rather than as a standalone technical solution.

Keywords – Learning analytics, predictive models, predictive analytics, student success, higher education, at-risk students

INTRODUCTION

Learning Analytics has become a central approach in higher education for understanding how higher education students exposed in increasingly digital and blended learning environments. By leveraging analytics from Learning Management Systems (LMS), Online assessments, and educational applications; institutions seek to monitor engagement, support self-regulated learning (SRL) and intervene before students fail, drop or withdraw. Recent systematic reviews shows that LA has been widely used to rapture SRL indicators such as time management, engagement and regularity of study; Which are strongly related to academic performance and persistence in higher education (Alhazbi et al., 2024). At the same time, Higher education student retention and drop-out remain major global challenges, with institution under

pressure to improve graduation rates and make better use of their growing educational data (Shafiq et al., 2022).

Current LA Research covers a broad landscape of techniques and contexts. Systematic reviews on student retention and dropout report extensive use of education data mining, machine leaning and deep learning models to identify at-risk students across traditional blended and fully online courses. Work on predictive modeling and meta-analysis shows that common behavioral indicators such as total online activity, forum participation and regularity of study are frequently used to predict academic success, through their predictive power can vary by course or design. Recent studies also explore interpretable models and explainable AI, for instance using Shapley Additive



explanations (SHAP) values to understand how engagement features such as virtual learning environment (VLE) clicks influence predictions of success (Oyedotun et al., 2025). Beyond predictive models, systematic models, systematic reviews have examined LA dashboards reporting that dashboards can increase participation but show limited and mixed evidence of improving achievements and motivation, especially when evaluation are small-scale or methodologically weak (Kaliisa et al., 2024). In parallel, LA has been applied in educational games, process mining and institutional innovation to support formative assessment, adaptive feedback and process monitoring.

Despite this progress, the literature highlights several important issues and limitations. Many SRL-related LA indicators are derived from time-on-task and engagement traces, yet are rarely validated against SRL theories, raising concerns about whether the measures truly represent self-regulation processes (Alhazbi et al., 2024). Empirical synthesis of predictive modes shows that behavior-based features often explain only a modest portion of variance in final grades, and that internal learner conditions (Such as motivation and prior knowledge) remain under-represented in trace models. Reviews of dashboards interventions report small or negligible effects on achievements and motivation, with frequent methodological problems such as self-selection bias, lack of control groups and the assumption that access equals actual use (Kaliisa et al., 2024). Moreover, meta-analytic and multi-course studies show that predictors found in one course or discipline often fail to replicate in others, limiting the portability and scalability of existing models.

There is also conceptual concept, ethical and innovation-related gaps in how LA is currently deployed. Empirical reviews of ethics in LA indicate that most studies focus on institutional rather than student perspectives and concentrate on issues such as transparency, privacy, and informed consent, while

giving less attention to justice, bias and student's agency in interpreting their own data (Cerratto Pargman & McGrath, 2021). Mapping studies of LA for educational innovation show that many projects emphasize technical infrastructures and process monitoring, but fewer integrate robust learner-centered design, strong theoretical ground or systematic evidence of impact in learning outcomes (Okoye et al., 2020). At the same time reviews of LA in educational games and interactive environments report high implementation complexity, heterogeneous metrics an a lack of generalize design guidelines, making it difficult for educators to adopt LA in a consistent and pedagogically meaningful way (Tlili, A., et al. 2021). These issues point to need for research that combines sound predictive modeling with attention to theory, ethics and practical usefulness for students and teachers.

Given the limitation, there is a clear need for further studies that develop and evaluate integrated, interpretable leaning analytics approaches for supporting student success and retention in higher education. Existing review calls for models that combine behavioral engagement indicators, SRL-related features and context-specific variables, while being grounded in learning theories and validate across multiple courses or cohorts. Recent work on explainable predictive modeling suggest that interpretable analytics could help institutions design fairer early warning system and make interventions more transparent to students and staff (Oyedotun et al., 2025). However, there is still limited empirical work that brings together student-centered design, ethical considerations and rigorous evaluation of the actual impact of LA tools (Such as dashboards or alerts) on retention and academic achievements. This gap motivates an in-depth investigation of how learning analytics and predictive modeling can be used not only to predict risk but also to meaningfully support students' ongoing learning and decision-making.

In light of this, the present study aims to investigate how learning analytics and predictive modeling can enhance student success and retention within the higher education. Specifically, the study will (1) identify and compare key behavioral, academic and SRL-related indicators that predict academic success and dropout risk; (2) develop an interpretable predictive model that uses these indicators to detect students at risk of failure or withdrawal; and (3) examine how such analytics can be embedded into student-facing tools (e.g., dashboard or feedback mechanism) in ways that are pedagogically grounded and ethically responsible. Building on prior work on retention analytics, LA for dropout prevention, and meta-analytic identification of robust predictors, the study seeks to provide both empirical evidence and practical design implications for institutions wishing to deploy LA in support of student success.

To guide this work, the study will address the following research questions, derived from the focal concerns highlighted in the reviewed literature:

1. Which learning analytic indicators (such as engagement traces, regularity of study assessment performance and SRL-related behavior) are the strongest predictors of student success and dropout risk in higher education courses?
2. How accurately and consistently can an interpretable predictive model identify at-risk students across different course offering or learning context?
3. In what ways can the output of such predictive models be integrated into student-centered analytics tools (such as dashboards or early-warning feedback) that are pedagogically meaningful, ethically sound and supportive of self-regulated learning and retention?

METHODOLOGY

2.1 General database search

The study employed a structured literature search to identify empirical and systematic works on learning analytics, predictive modeling and student success in higher education. Google Scholar served as the main search engine, from which full text were accessed through linked publisher and journal platforms, including *Wiley online library* for studies on self-regulated learning analytics [1] *The Journal of Learning Analytics* via the *Diva portal for ethics-related work* [2] *ResearchGate* for hosted copies of systematic reviews on retention and predictive analysis [3] *MDPI* for dropout prevention studies [4] *ScienceDirect* for research on predictors on academic success [5] the *ACM Digital Library* website such as *Journal of Science Research and Reviews* [7] *Studies in Higher Education* [8] *International Journal of Interactive Media and Artificial Intelligence* [9] and *International Journal of Computer Information System and Industrial Management Applications* [10] The search strategy used combinations of the following keywords and phrases: *Learning analytics, predictive models, predictive analytics, student success, higher education, at-risk students*; with Boolean operators (E.g. “*Learning analytics*” AND “*Student Retention*” Etc.) to focus the results on studies addressing early identification of at-risk students , academic performance and retention.

Information from retrieved studies was systematically collected and organized into a summary table. For each article that met the inclusion criteria the authors, year of publication, objectives of the study (E.g. measuring self-regulated learning, predicting academic success, preventing dropout, mapping ethics or early indicators), educational setting (with explicit focus o higher education institutions), and methodological approach (such as systematic literature review, meta-analysis, predictive modeling, dashboard, evaluation, or mapping study)were recorded. The selection was limited to works published between 2020

and 2025, written in English, that (a) Focused on learning analytics or educational data mining, (b) addressed predictive or explanatory models related to student success, at risk identification, or retention in higher education, and (c) provided empirical data or systematic synthesis rather than purely conceptual discussion. Studies were excluded if they did not involve higher education, did not use learning analytics, or predictive/educational data mining methods, or lacked sufficient methodological details. Applying these criteria yielded a final list of 10 studies that collectively inform the present work's focus on leveraging analytics and predictive modeling to support student success and retention (Alhazbi et al., 2024; Okoye et al., 2020).

2.2 Identifying the issues and challenges in Predictive Learning Analytics for Student Success in Higher Education

The researcher analyzed the selected articles to identify the main issues and challenges in predictive learning analytics for student success in higher education. This step was based on different kinds of studies in reference list. Articles or references are important because it shows how predictive learning analytics is currently used and designed. It also reveals repeated methodological, ethical, and practical problems. For this work, the literature was first grouped with strong practical evidence, where report empirical predictive models, systematic reviews or meta-analyses on retention, dropout, academic success or early indicators in higher education. The second group are those with conceptual ideas, those are kept as a source of ideas and design directions for later analysis. Articles that did not clearly related to predictive learning analytics, higher education or student success and at-risk identification were excluded.

2.3 Recognizing the Solutions for the Issues and Challenges in Predictive Learning Analytics for Student Success in Higher Education

After the issues were identified, the researcher looked for solutions and future directions by re-reading the selected articles and checking for links and shared ideas across them. In the first part of analysis, common challenges, such as narrow or weakly validated indicator of self-regulated learning (Alhazbi et al., 2024). Limited transferability of behavioral predictors across courses and context (de Oliveira et al., 2021; Saqr et al., 2022), ethical concerns about transparency, consent and student outcomes (Kaliisa et al., 2024) and incomplete framework for early indicators and retention strategies (Shafiq et al., 2022; de Oliveira et al., 2021; Okoye et al., 2020) were matched with solutions suggested in literature. These solutions include using more theory-based, multi-dimensional predictive models that combine behavioral traces, SRL indicators and context factors (Alhazbi et al., 2024; Jovanović et al., 2021; Oyedotun et al., 2025; Saqr et al., 2022; Okoye et al., 2020) applying systematic and data-driven approaches to retention and early warning (Shafiq et al., 2022; de Oliveira et al., 2021; Oyedotun et al., 2025) using meta-analytic and mapping methods to identify stable predictors and success factors across different studies and institutions (Saqr et al., 2022; Okoye et al., 2020) and designing analytic tool, dashboard and game-based environment that builds in ethical principles and student-centered design from the start (Cerratto Pargman & McGrath, 2021; Kaliisa et al., 2024; Tlili et al., 2021). In the second part, these connected ideas were turned into anticipated resolutions to guide the present study. For instance, focusing on interpretable and reusable predictive models, planning validations across multiple courses, and ensuring that predictive information is presented to students and staff in ways that are ethically sound and pedagogically helpful so that, later stages of the research directly address the issues found in the existing literature (Alhazbi et al., 2024; Okoye et al., 2020).

2.4 Identify factors of remaining challenges affecting Predictive Learning Analytics for Student Success in Higher Education.

The challenges identified by the researcher based on the literature includes with model probability across different courses and discipline, as well as the underrepresentation of socio-emotional factor such as motivation and self-regulation. Ethical concerns, such as bias in demographic, data and lack of transparency in models, also persist. Data imbalance, overfitting, and the limited impact of leaning analytics dashboards (LADs) on student outcomes further complicates the effectiveness of LA. Additionally, many models lack validations across diverse learning context, limiting their generalizability. These factors highlight the need for future research to address these challenges and improve the impact of LA.

RESULTS AND DISCUSSION

3.1 General database search

Table 1 shows the number of papers identified in each of the databases using the key search query *Predictive Learning Analytics for Student Success in Higher Education* between 2020 until 2025. The first column shows the total number of results retrieved, including both peer-reviewed papers and other non-scientific publications such as magazine articles. In total, 125 articles were initially examined across 9 different databases.

Based on a careful consideration and review of abstracts, 50 papers were excluded from further consideration; as did not meet the inclusion criteria (e.g., non-peer-reviewed, irrelevant to the research focus, or; lack of empirical data) The remaining 75 papers were assessed for their relevance to the research topic. Of these, 65 papers were excluded after a full-text review, as they did not provide contributions to the research gap in Predictive Learning Analytics or were methodologically unsuitable.

Ultimately, 10 peer-reviewed research paper on *Predictive Learning Analytics for Student Success in Higher Education* were selected for review. These papers, all published in reputable journals and

conferences, represent the most recent and relevant studies in the field. They provided strong empirical evidence and directly addressed the identified research gap, forming the foundation of this literature review.

Table 1: Result from searched databases.

Database	Total Number of Result	Peer Review Paper
Wiley's Journal of Computer Assisted Learning	8	1
SLAR Journal of Leaning Analytics	9	1
IEEE Xplore	11	1
MDPI	3	1
Elsevier's Computers and Education	25	1
ACM Digital Library	8	1
Journal of Science Research and Reviews	2	1
Taylor and Francis's Studies in Higher Education	9	1
UNIR International Journal of Interactive Multimedia and Artificial Intelligence	18	1
Scopus	14	1
Google Scholar	17,000	N/A

3.2 Issues and challenges in identifying Predictive Learning Analytics for Student Success in Higher Education.

Across reading peer-reviewed studies or literature reviewed, the evidence suggest that predictive leaning analytics (PLA) for student success is till constrained by recurring development and implementation issues that complicate robust predictors and models are, identified in higher education setting. A consistent theme is that many analytics indicators are drawn from digital traces such as LMS activity, timing,

clicks, and participation; yet the construct validity of these proxies remains difficult to confirm, particularly when indicators are used to represent complex leaning process such as self-regulated learning, without strong validation against SRL theory or direct measures (Alhazbi et al., 2024).

This challenge affects identification because predictors can appear statistically useful while still providing an incomplete or potentially misleading representation of learning. Then, closely related challenges concerns model portability and generalizability. Multiple studies emphasize that predictors of academic success do not consistently replicate across different learning designs, subject areas, cohorts or institutions, which limits confidence in applying “Known predictors” beyond the original context (Saqr et al., 2022). Even within relatively similar course environments, variability across iterations and across courses can remain moderate, and some commonly used indicators may be statistically insignificant or less portable across offering (Saqr et al., 2022). In systematic review evidence, lack of generalizability is attributed to design that focus on single course or narrow settings, where assessment structures and marking practices differ substantially, making cross-course identification of stable predictors difficult (Shafiq et al., 2022).

Methodological and data constraints further complicate the identification of PLA models that are dependable for student success monitoring. Reviews highlight common risk such as data imbalances and overfitting, which can bias patterns and weaken the trustworthiness of report model performance (Shafiq et al., 2022). These issues are often intensified by limited sample sizes, where datasets below typical large-scale thresholds reduce the reliability of derived conclusion and make it harder to confirm whether predictors are stable. At the same time, reliance on traditional performance variables, such as: grades, attendance, and demographic, can underrepresent personal, cognitive,

and behavioral conditions that may meaningfully shape success outcomes, thereby narrowing the predictor space and limiting explanatory adequacy. Ethical and governance concerns remain embedded in the identification problem, because predictive systems are developed and interpreted within institutional constraints that affect student agency and the legitimacy of analytic decisions.

Ethical review evidence frames leaning analytics as a complex socio-technical practice, where privacy and consent concerns coexist with broader issues linked to how analytics are designed, explained and acted upon in educational context (Cerratto Pargman & McGrath, 2021). Related to this, limitations in reporting and transparency can still appear at the study level. For instance, when key technical details like algorithm specifications or validation process; are not fully disclose, reducing reproducibility and weakening confidence in what counts as an “identified” best approaches.

Finally, the reviewed literature suggests that expanding PLA into richer leaning environments introduces additional identification complexity. In game-based and interactive context, the empirical base is still described as relatively limited, while evidence remains fragmented across diverse type and designs, making it harder to identify transferable design guidelines or consistent predictors that support student success at scale (Kaliisa et al., 2024).

Table 2: Challenges in Predictive Learning Analytics for Student Success in Higher Education.

Challenges	Peer-Reviewed Studies or Literature/References Reviewed
SRL Indicators are often trace-based proxies and	Alhazbi et al., 2024



not well validated against SRL Theory

Difficulty identifying "suitable observable indicators" from Large LMS trace datasets
Alhazbi et al., 2024

Overreliance on traditional /static predictors; Limited inclusion of broader learner-related factors.
Shafiq et al., 2022

Limited portability/ inconsistent predictors across courses and design
Jovanović et al., 2021, Saqr et al., 2022

Context dependence: Internal conditions and instructional context shape predictive relations
Jovanović et al., 2021

Data quality, imbalances, and overfitting threaten model robustness
Shafiq et al., 2022

Dropout is multifaceted with heterogeneous variables, complicating feature definition
De Oliveira et al., 2021

Dashboards show mixed /limited evidence on achievement/motivation ; evaluation design often weak.
Kaliisa et al., 2024

System-level obstacle; Information overload, incoherent metrics, misinterpretation risk; Privacy/ethics constrains.
Okoye et al., 2020

Ethics emphasis on transparency/privacy/ consent with less attention to justice, bias, student agency
Cerratto Pargman & McGrath, 2021

Need for interpretability/ explainable outputs to support responsible use
Oyedotun et al., 2025

3.3 Identified solutions for issues and challenges in Predictive Learning Analytics for Student Success in Higher Education

Based on the peer-reviewed paper/references, the proposed solutions are mainly concerned with how predictive learning analytics (PLA) should be developed and applied, so that it can meaningfully support student success in higher education, rather than remaining as isolated technical models. Across the literature, the suggested directions for improvement focus on related areas: First, strengthening how predictive indicators are defined and validated; Second, improving model design and testing; so that predictions remain reliable across different courses and cohorts; and third, ensuring that predictive outputs are implemented in ways that are interpretable, ethically responsive.

3.3.1 Improve indicator quality through theory-based definition and multi-source evidence

A consistent solution across the reviewed studies is to treat indicator selection as a learning measurement issue, not only a technical-selection task. In particular, SRL-related indicators are frequently

derived from trace data such as engagement, regularity, and time-management patterns, yet these proxies may not represent SRL processes unless they are clearly grounded in SRL theory and validated against strong evidence (Alhazbi et al., 2024). Therefore, the literature supports defining indicators through established SRL framework and avoiding the assumption that “more clicks” or more “time online” automatically indicates effective leaning (Alhazbi et al., 2024).

3.3.2 Develop context-sensitive models

To address limited portability and weak generalizability, the reviewed literature emphasizes that predictive relationship changes across context because student success is shaped by both internal leaning conditions and instructional conditions. As a result, the same behavioral indicators may carry different meaning in different course designs, assessment structures or delivery modes (Jovanović et al., 2021). A key solution is therefore to design PLA models that explicitly account for course context and to avoid assuming that “known predictors” will transfer directly across disciplines or cohorts (Jovanović et al., 2021; Saqr et al., 2022).

Evidence from multi-course and synthesis work reinforces that predictor importance varies across offering, which makes cross-context validation a management requirement rather than an optional technical step (Saqr et al., 2022). In practice, this implies that institution should validate models across multiple course iterations and cohorts before scaling and should document the boundaries of model use such as where the model performs well and where it does not (Jovanović et al., 2021; Saqr et al., 2022).

For dropout prediction, the literature further highlights that dropout is multifaceted and influenced by heterogeneous variables, suggesting that context-aware modeling and prevention requires broader designs rather than single-pattern explanations based only on online engagement (De Oliveira et al., 2021).

3.3.3 Strengthen robustness through data practices, imbalance handling, and rigorous evaluation designs

Several studies frame predictive reliability as strong dependent on dataset quality and validation design. Retention-focused synthesis identifies that producing high-quality data and selecting influential factor is one of the most difficult parts of predictive development, and that risk such as data imbalance and overfitting can weaken reported model performance if not carefully addressed (Shafiq et al., 2022). Accordingly, the reviewed literature supports strengthening data practices through clearer preprocessing procedures, explicit handling of imbalanced.

In addition, the literature suggest that predictive usefulness should be judged not only by accuracy, but also by implications of false positives, and false negatives in early warning system, because these errors directly affect who receives support and who may be overlooked (Oyedotun et al., 2025). As a result, robust PLA implementation requires transparent reporting of modeling decisions, which includes validation strategies; and performing monitoring across cohorts, so institutions can assess whether models remain stable and reliable over time (Shafiq et al., 2022; Oyedotun et al., 2025).

3.3.4 Embed interpretability and ethical governance into model development and deployment

Another clear direction is that PLA must be interpretable and ethically governed, to be credible in higher education. Ethical synthesis highlights that leaning analytics research frequently emphasizes transparency, privacy, and informed consent, while giving less attention to justice, bias, and student agency (Cerratto Pargman & McGrath, 2021). This implies that the PLA should be developed with explicit governance principles: students should be treated as stakeholders, predictive use should be explained, and institutions should be cautious about how risk labels are

communicated and acted upon (Cerratto Pargman & McGrath, 2021).

In parallel, studies on explainable modeling show that interpretability methods such as SHAP; can help clarify how specific indicators contribute to predictions, supporting transparency and trust among staff and students (Oyedotun et al., 2025). However, the same evidence also indicates that fairness claims require strong evaluation, reinforcing the need for unbiased auditing and careful oversight when demographic or sensitive variables are presented or when predictions may influence educational opportunities (Cerratto Pargman & McGrath, 2021; Oyedotun et al., 2025). These ethical and interpretability priorities also align with mapping evidence that warns about misinterpretation risk and information overload in real deployments, indicating that responsible governance is closely linked to usability and institutional feasibility (Okoye et al., 2020).

3.3.5 Connect prediction to student success through interventions and stronger tool evaluation.

The reviewed literature/references suggest that identifying risk is not sufficient unless PLA outputs lead to meaningful education actions. Systematic review evidence on leaning analytics dashboards indicates that dashboards may increase participation, but evidence for consistent improvement in achievement and motivation remains limited, and evaluation are often methodologically weak (Kaliisa et al., 2024). Therefore, literatures/references support treating dashboards, alerts, and early warning outputs as intervention tools, that must be designed for actionability and evaluated for impact, rather than treated as passive reporting system (Kaliisa et al., 2024).

Dropout-focused reviews further indicates that prevention requires connecting predictive identification to practical support mechanisms, such as academic

assistance and improved social involvement, reinforcing that PLA should be integrated into student support workflow, instead of being used only for classification (De Oliveira et al., 2021). Mapping work similarly emphasizes that educational innovation through leaning analytics requires clear objectives, credible implementation approaches and attention to the practical challenges that limit uptake and effectiveness.

3.4 Recognizing the factors that remain challenges affecting Predictive Learning Analytics for Student Success in Higher Education.

Based on the peer-reviewed papers/literature/references, the remaining challenges affecting predictive learning analytics (PLA) for student success in higher education can be recognized as several connected factors that still limit model reliability, transferability and practical value. First, indicators validity remains a central concern because many PLA studies rely on LMS trace data, such as click, logins, timing and participation, as proxies for learning. However, the papers/Literature/references indicates that these proxies are often weakly validated when they are used to represent complex learning process such as self-regulated leaning (SRL), which creates a risk that models predict activity rather than meaningful learning behavior (Alhazbi et al., 2024).

At the same time, context dependence continues to constrains portability and generalizability, since predictors of success and risk do not consistently replicate across different courses, disciplines, cohorts, and institutions. Predictive relations may shift when course design, assessment, structures or instructional conditions change, which makes it difficult to apply “known predictors” confidently beyond the original setting (Jovanović et al., 2021; Saqr et al., 2022). Systematic review evidence further links this limitation to narrow research design, where single-course studies reduce the ability to identify stable predictors that can transfer across context at scale (Shafiq et al., 2022).



Moreover, methodological and data-related constraints remain influential. It highlights ongoing risks such as data imbalance and overfitting, particularly in dropout or at-risk predictors when negative outcomes may be less frequent, and this risk can distort reported performance if not carefully managed (Shafiq et al., 2022). These concerns are often strengthened by small datasets and limited validation across cohorts, which reduces confidence that predictive models remain stable over time (Shafiq et al., 2022). In addition, several studies suggest that predictive approaches can become overly dependent on traditional variables such as grades, attendance, and demographic data; while broader leaning conditions that shape success may be underrepresented, narrowing the predictor space and limiting explanatory strength (Shafiq et al., 2022; Jovanović et al., 2021).

In parallel, interpretability and transparency remain important factors because PLA outputs are used to guide educational decisions. Even where explainable methods are introduced, it emphasizes that prediction must be judged in relation to misclassification risk and supported by clear reporting modeling and validation procedures to strengthen trust and reproducibility (Oyedotun et al., 2025). There are also ethical and governance-related concerns that remain embedded in PLA practice. Ethical review evidence shows strong attention to transparency, privacy, and consent, while issues linked to justice, bias, and student agency are less consistently addressed, even though they shape the legitimacy of predictive system and how students experience analytics-driven decisions (Cerratto Pargman & McGrath, 2021).

Finally, implementation and impact challenges continue to affect PLA effectiveness. Studies on dashboards report mixed evidence for improvement in achievement and motivation and highlight common evaluation limitations such as self-selection and the assumption that access implies meaningful use (Kaliisa et al., 2024). Mapping further evidence points to

practical barriers such as information overload, heterogeneous metrics, and misinterpretation risk, which reduces educational usefulness if PLA is not aligned with clear objectives and support process (Okoye et al., 2020). Similarly, in more complex environment such as game-based settings, implementation complexity and fragmented evidence make it difficult to establish transferable guidance for adoption scale (Tlili et al., 2021).

Overall, the reviewed papers/Literature/references suggest that these factors are interrelated, and addressing them requires coordinated attention to measurement validity, contextual design, methodological robustness, ethical governance, and implementation practices in order for PLA to support student success in higher education (Alhazbi et al., 2024; Okoye et al., 2020).

CONCLUSION

This work synthesized peer-reviewed studies/literature that examine Predictive Learning Analytics (PLA) for student success in higher education. Across these papers, PLA is primarily operationalized through indicators derived from digital leaning environments such as LMS trace data; and academic performance records, with the goal of identifying at-risk students early enough to support retention and achievement. However, the reviewed evidence suggest that PLA effectiveness depends less on the availability of data and more on how well indicators are defined, validated and interpreted in relation to learning process they claim to represent. In particular, SRL-related indicators are often implemented as trace-based proxies, and the literature cautions that these measures can be difficult to justify without theoretical alignment and validation. Overall, the review shows that several limitations remain persistent across the field. Predictive relationships frequently change across courses and contexts, reducing confidence that “known predictors” or models will perform consistently when transferred beyond the



original setting. Methodological contains; including data quality issues, imbalance, and risks of overfitting also continue to affect the credibility and reproducibility of reported model performance. In addition, it emphasizes that PLA is not only a technical activity. Instead, it covers ethical consideration such privacy, consent, transparency, and student agency; shape whether predictive systems are trusted and responsibly used, while evidence on dashboard-type implementations remains mixed and often limited by weak evaluation design.

FUTURE WORK

Future researcher should extend the present work by addressing additional questions that remains central to the Predictive Learning Analytics (PLA) for students' success in higher education. However, due to study's scope and time constrains, the researcher synthesizes and focused on 10 peer-review papers/literature/paper, as the researcher saw relevant data needed to answer his research questions. In particular, future researchers should examine the stability of predictors and model performance across diverse institutional settings, disciplines, cohorts, and learning design, given continuing concerns about portability and generalizability in existing studies.

Additional work is also needed to strengthen how self-regulated leaning (SRL) and other learner conditions are operationalized in PLA, since many commonly used trace-based indicators are not consistently validated against the SRL theory and may capture activity rather than learning process. Moreover, future studies should investigate interpretability requirements for different stakeholders such as: Students, instructors/lecturers, advisors; especially because misclassification in early-warning system has practical consequences for support allocation and student outcomes. At the same time, stronger empirical evaluation is required to determine whether dashboards and early-warning tools produce measurable improvements in achievement, motivation, and

retention as current evidence remains mixed and is often limited by methodological weakness. Finally, Future researcher should research further on ethical governance and operational student agencies, particularly around transparency, privacy, consent, and potential bias. Thus, should examine the institutional implementation condition that shape responsible and sustainable adaptation, including risk of misinterpretation, information and overload, and heterogeneous indicator practices.

REFERENCES

- Alhazbi, S., Al-Ali, A., Tabassum, A., Al-Ali, A., Al-Emadi, A., Khattab, T., & Hasan, M. A. (2024). Using learning analytics to measure self-regulated learning: A systematic review of empirical studies in higher education. *Journal of Computer Assisted Learning*, 40(4), 1658–1674.
- Cerratto Pargman, T., & McGrath, C. (2021). Mapping the ethics of learning analytics in higher education: A systematic literature review of empirical research. *Journal of Learning Analytics*, 8(2), 123–139.
- Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2022). Student retention using educational data mining and predictive analytics: A systematic literature review. *IEEE Access*, 10, 72480–72494.
- de Oliveira, C. F., Sobral, S. R., Ferreira, M. J., & Moreira, F. (2021). How does learning analytics contribute to prevent students' dropout in higher education: A systematic literature review. *Big Data and Cognitive Computing*, 5(4), Article 64.
- Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education*, 172, Article 104251.



- Kaliisa, R., Misiejuk, K., López-Pernas, S., Khalil, M., & Saqr, M. (2024). Have learning analytics dashboards lived up to the hype? A systematic review of impact on students' achievement, motivation, participation and attitude. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 295–304).
- Oyedotun, S. A., Ejenarhome, O. P., & Oise, G. P. (2025). Learning analytics and predictive modeling: Enhancing student success through data-driven insights. *Journal of Science and Reviews*, 2(3), 42–51.
- Saqr, M., Jovanovic, J., Viberg, O., & Gašević, D. (2022). Is there order in the mess? A single paper meta-analysis approach to identification of predictors of success in learning analytics. *Studies in Higher Education*, 47(12), 2370–2391.
- Tlili, A., et al. (2021). A systematic literature review of empirical studies on learning analytics in educational games. *International Journal of Interactive Multimedia and Artificial Intelligence*, 7(2), 250–262.
- Okoye, K., Nganji, J. T., & Hosseini, S. (2020). Learning analytics for educational innovation: A systematic mapping study of early indicators and success factors. *International Journal of Computer Information Systems and Industrial Management Applications*, 12, 138–154.