US JOURNAL OF NEW INSIGHTS IN TECH & EDUCATION

Vol. 3, No. 2 · Article ID: USJNITE-2304 · Journal Homepage: www.edtechinsights.org

Early Identification of At-Risk Students Using Learning Analytics: A Data-Driven Approach to Supporting Academic Success

MD Shahadat Hossain Shishir[1], Md Rasel Ul Alam [2], Kazi Arham Ahmed [3]

[1] *World University of Bangladesh, Dhaka-1230*

[2] *University of the Cumberlands, Kentucky, USA*

[3] *University Kebangsaan Malaysia, Selangor, Malaysia*

**Corresponding author: shahadatdaf20@gmail.com*

Received: 20 July 2023; Revised: 29 August 2023; Accepted: 26 September 2023; Published: 06 November 2023

KEYWORDS

Learning analytics
At-risk students
Early warning systems
Predictive analytics
Educational data mining
Student success
Academic retention
Machine learning

Abstract

Learning analytics has created new opportunities to identify students at risk of academic failure, withdrawal, or disengagement before negative outcomes become difficult to reverse. This paper synthesizes research on the use of learning-management-system activity, assessment records, administrative data, and other educational traces to support early identification and timely academic intervention. The synthesis examines data sources, predictive techniques, risk-detection timing, and the translation of predictions into actionable support. The findings indicate that effective early-warning approaches combine behavioral and academic indicators rather than relying on a single measure. Common signals include declining learning activity, weak assessment performance, late or missing submissions, irregular participation, and prior academic risk. Predictive approaches include logistic regression, decision trees, random forests, k-nearest neighbors, support-vector methods, neural models, ensemble techniques, and time-series analysis; however, predictive accuracy alone does not establish educational value. A key challenge is translating prediction into effective intervention, as identifying high-risk students does not necessarily ensure improved outcomes. The paper therefore proposes a data-to-action framework linking early identification with human review, transparent risk communication, targeted outreach, and monitored support. Privacy, fairness, transparency, false positives, and responsible use of student data are treated as central design requirements. The study concludes that learning analytics is most educationally valuable when predictive capability is embedded within a responsive support system that enables institutions to act early, appropriately, and transparently.

1. Introduction

Higher education institutions increasingly generate detailed digital traces of learning activity through learning management systems (LMSs), assessment platforms, attendance systems, student-information systems, discussion environments, and other educational technologies. These traces make it possible to observe patterns of participation at a granularity that was difficult to achieve through conventional end-of-term assessment alone. The central promise of learning analytics is therefore not simply to describe what students have done, but to use timely evidence to understand what may happen next and to support decisions that improve learning and persistence (Ifenthaler & Yau, 2020; Namoun & Alshanqiti, 2021).

The idea of an early-warning system emerged from this shift. Macfadyen and Dawson (2010) demonstrated that LMS data could be mined to identify students who might require attention, establishing an influential proof of concept. Subsequent research extended this approach through institutional early-alert systems, predictive classification, clickstream analysis, dropout modeling, and machine-learning methods (Arnold & Pistilli, 2012; Jayaprakash et al., 2014; Akçapınar et al., 2019). The field has consequently moved from asking whether student risk can be predicted to asking how early, how accurately, and under what conditions predictive information can be converted into effective support.

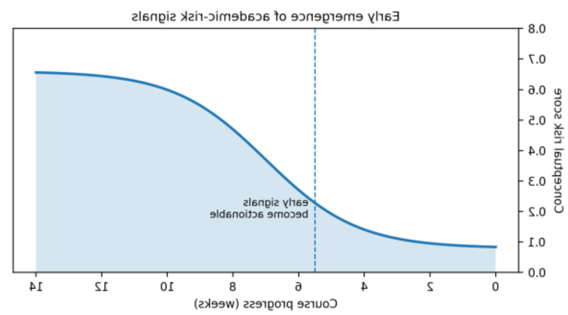

Figure 1. Conceptual emergence of academic-risk signals over the course timeline.

An at-risk student is not a fixed type of learner. Risk is better understood as a time-dependent probability or condition arising from the interaction of prior achievement, current engagement, learning behavior, course demands, institutional context, and other factors. Studies such as Marbouti et al. (2016), Howard et al. (2018), and Bañeres et al. (2019) illustrate how predictive models can detect risk before a course or academic period concludes. In distance and online education, the value of earlier detection can be particularly important because reduced physical visibility may make changes in participation harder for instructors to notice through informal observation alone.

Yet prediction is not equivalent to support. Jayaprakash et al. (2014) examined both predictive performance and the challenges of interventions, while Dawson et al. (2017) evaluated a retention-oriented learning analytics program and highlighted the distinction between identifying students and producing measurable impact. This distinction defines an important conceptual problem for learning analytics research: a model can be statistically successful while an early-warning program remains educationally ineffective if alerts are too late, too opaque, too frequent, poorly interpreted, or not followed by appropriate human action.

This paper addresses that problem by synthesizing research around a complete data-to-action pathway: data capture, feature engineering, risk prediction, early classification, human interpretation, intervention, and academic-support outcomes. The objective is not to propose another isolated machine-learning model, but to identify design principles that can help institutions use predictive learning analytics responsibly and productively.

1.1 Problem statement

The literature contains substantial evidence that academic risk can be predicted from educational data, yet studies differ widely in data sources, target outcomes, prediction horizons, modeling approaches, evaluation metrics, and intervention designs (Cui, 2019; Namoun & Alshanqiti, 2021). This heterogeneity makes it difficult to identify a single universally superior predictive approach. More importantly, research has been more developed on the technical question of prediction than on the organizational question of what happens after a student is flagged. A useful framework must therefore connect predictive accuracy with timeliness, interpretability, human review, and actionable support.

1.2 Research objectives

I. Identify the principal educational data sources and behavioral indicators used to identify at-risk students.
II. Synthesize the predictive and learning-analytics techniques used for early-risk classification.
III. Examine how the timing of predictions influences their potential usefulness for academic support.
IV. Clarify the gap between risk prediction and successful intervention.
V. Propose an integrated data-to-action framework for responsible early identification.

1.3 Research questions

I. RQ1. Which data sources and learning behaviors are most frequently used to identify students at academic risk?
II. RQ2. Which predictive approaches are used to classify or forecast academic risk?
III. RQ3. How does prediction timing affect the usefulness of early-warning information?
IV. RQ4. What factors influence the transition from accurate prediction to effective student support?

2. Literature Review

2.1 Learning analytics and the logic of early identification

Learning analytics can be broadly understood as the measurement, collection, analysis, and reporting of data about learners and learning contexts for purposes of understanding and optimizing learning (Ifenthaler & Yau, 2020). In early-identification research, analytics transforms routine educational traces into indicators of changing engagement or performance. This allows institutions to intervene before a final grade, withdrawal, or academic exclusion has already occurred. Macfadyen and Dawson (2010) established the practical basis for this logic using LMS activity. Jayaprakash et al. (2014) expanded it into a scalable early-alert initiative that combined risk prediction with institutional processes for responding to students.

2.2 Defining academic risk

The operational definition of 'at risk' varies across studies. Some research predicts course failure or unsuccessful completion (Marbouti et al., 2016; Zacharis, 2015), some predicts dropout or withdrawal (Berens et al., 2019; Olaya et al., 2020), and some focuses on final examination performance or course grades (Chen & Cui, 2020; Musso et al., 2020). Ifenthaler and Yau (2020) note that study success can range from completing a degree to completing individual learning tasks. This variation matters because a model optimized for final-grade prediction may not be the same model needed to identify students who are beginning to disengage several weeks before the end of a course.

2.3 Data sources and feature construction

The literature relies heavily on LMS and clickstream activity because these sources provide repeated observations of student behavior. Common features include frequency of logins, resource views, time spent in learning environments, discussion activity, submission patterns, and assessment attempts (Macfadyen & Dawson, 2010; Casey & Azcona, 2017; Aljohani et al., 2019). Assessment and grade history provide a different form of signal because they represent demonstrated academic performance. Administrative variables may add prior achievement, program characteristics, or enrollment context (Berens et al., 2019). The strongest conceptual argument emerging from the literature is therefore for multimodal rather than single-source prediction: behavior can indicate changes in engagement, whereas grades and prior records provide evidence of demonstrated attainment.

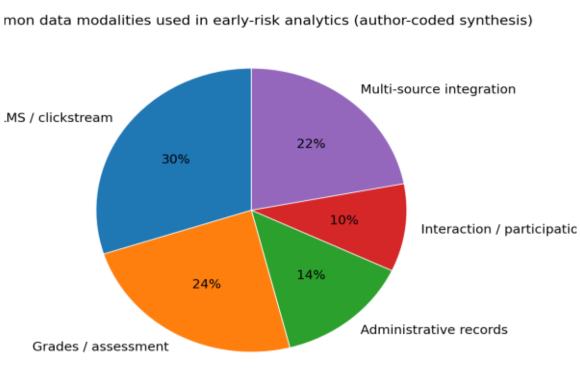

Figure 2. Common data modalities used in early-risk analytics.

2.4 Predictive modeling approaches

A wide range of statistical and machine-learning approaches appear in the literature. Logistic regression and related interpretable classification methods remain important because institutions often need to understand why a student was flagged. Decision trees, random forests, k-nearest neighbors, support-vector approaches, neural models, and ensemble methods are also used when nonlinear patterns or interactions are expected (Howard et al., 2018; Namoun & Alshanqiti, 2021). Akçapınar et al. (2019), for example, compared algorithms and preprocessing strategies and showed that early course interaction data could be used to predict unsuccessful performance. Karalar et al. (2021) similarly applied an ensemble approach to academic-failure prediction in a distance-learning context.

The literature therefore suggests that model selection should be driven by the intended educational use rather than accuracy alone. A marginally more accurate but opaque model may be less useful for advisors than a slightly less accurate model whose signals can be explained and acted upon. Interpretability, calibration, false-positive costs, prediction horizon, portability, and computational burden are consequently part of the model-selection problem.

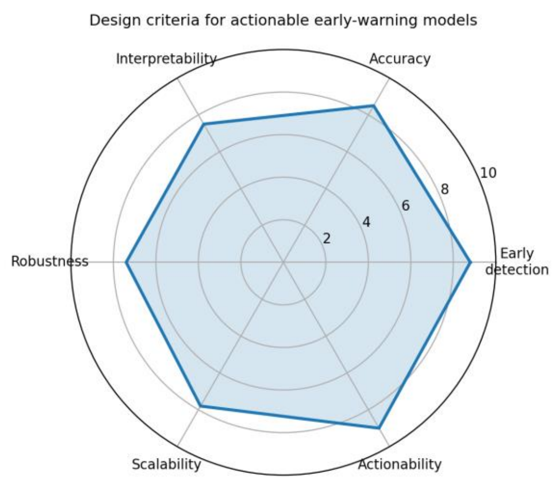

Figure 3. Design criteria for an actionable early-warning model.

2.5 Prediction timing and early-warning horizons

Early identification is valuable only when it creates sufficient time for meaningful action. Wolff et al. (2014) explicitly investigated early detection in distance-learning modules. Marbouti et al. (2016) examined prediction across course progression, while Chen and Cui (2020) used time-series LMS behavior for early course-performance prediction. These studies support a central principle: the prediction window must be selected with the intervention window in mind. A highly accurate prediction made after most learning opportunities have passed can have less practical value than a moderately accurate signal made early enough to trigger constructive support.

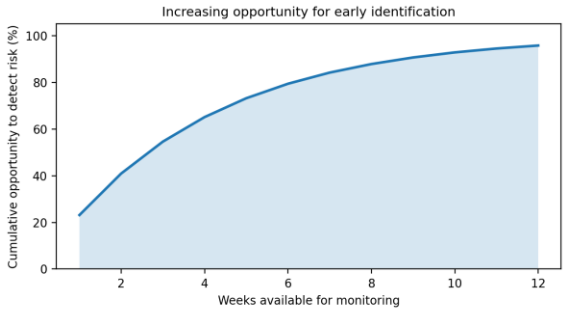

Figure 4. Increasing opportunity for early identification as more course weeks become observable.

2.6 Early-warning systems and institutional implementation

Arnold and Pistilli's (2012) Course Signals initiative is an influential institutional example of using learning analytics to support student success. Jayaprakash et al. (2014) similarly addressed scalability, portability, and intervention within an open academic analytics initiative. These studies show that an early-warning system is more than a classifier. It requires data pipelines, thresholds, communication mechanisms, personnel, and an institutional response process. Waddington et al. (2016) further showed that categorized course-resource usage can improve early-warning systems, illustrating the importance of feature design and contextual interpretation.

2.7 From prediction to intervention

One of the most important findings across the evidence base is that predictive success and intervention success must be evaluated separately. Dawson et al. (2017) examined a learning-analytics retention program rather than only a predictive model. Jayaprakash et al. (2014) described intervention as part of the broader early-alert process. Olaya et al. (2020) advanced the logic by studying uplift modeling for dropout prevention, an approach intended to distinguish students who may benefit from intervention from those for whom an intervention may have little effect. This is a critical conceptual move: the target is not merely 'who is likely to fail?' but 'who is likely to benefit from a specific support action?'

2.8 Ethical and governance considerations

Early-warning systems operate on sensitive educational data and can shape how students are perceived and treated. Privacy, informed consent, fairness, transparency, data governance, and the risk of stigmatizing students therefore require explicit consideration. Ifenthaler et al. (2021) argue for putting learning back into learning analytics, emphasizing responsible and educationally grounded use of data. Jones (2019) likewise highlights informed-consent mechanisms and student autonomy in higher-education learning analytics. A responsible system should minimize unnecessary data collection, protect records, communicate risk cautiously, preserve human judgment, and provide students with opportunities to understand and respond to support decisions.

2.9 Research gap

The literature provides strong evidence for predictive modeling, but it remains fragmented across different target outcomes, data environments, prediction windows, and institutional contexts. Systematic reviews by Ifenthaler and Yau (2020), Namoun and Alshanqiti (2021), Albreiki et al. (2021), and de Oliveira et al. (2021) demonstrate a mature technical literature, yet they also reveal substantial methodological heterogeneity. The most consequential gap is the limited integration of four elements in one design logic: early prediction, interpretation, intervention matching, and outcome evaluation. The contribution of this paper is therefore a data-to-action framework that treats predictive learning analytics as one component of a broader academic-support system.

Table 1: Core data sources and their analytical role

Data source	**Examples**	**Analytical purpose**	**Representative evidence**
LMS/clickstream	Logins, views, activity frequency, resource access	Detect changes in engagement	Macfadyen & Dawson, 2010; Casey & Azcona, 2017
Assessment data	Grades, quizzes, submissions, attempts	Measure demonstrated attainment	Marbouti et al., 2016; Chen & Cui, 2020
Administrative data	Prior performance, enrollment variables	Add historical/contextual signal	Berens et al., 2019; Musso et al., 2020
Multi-source data	al + academic + administrative data	Improve context and robustness	Ifenthaler & Yau, 2020; Albreiki et al., 2021

3. Conceptual Framework

The proposed framework treats early identification as a sequence of linked decisions rather than a single predictive task. First, educational systems collect longitudinal traces. Second, raw traces are transformed into pedagogically meaningful features. Third, a predictive model estimates risk or an outcome of interest. Fourth, risk scores are interpreted using thresholds, trends, and contextual information. Fifth, human staff determine whether an alert warrants outreach. Finally, support is delivered and its effect is monitored. This sequence reflects the literature's movement from prediction toward action and avoids equating a high-risk score with a predetermined student outcome.

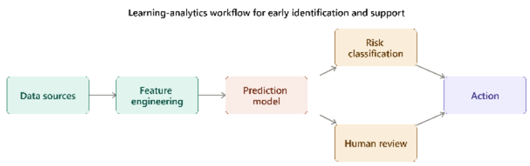

Figure 5. Learning-analytics workflow connecting data capture, prediction, human review, and support.

3.1 Data layer

The data layer contains repeated observations rather than isolated snapshots. This supports trend-sensitive features such as declining participation, increasing missed submissions, or repeated low assessment results. Time-series information is especially important for early-warning systems because the direction and pace of change may matter as much as absolute values (Chen & Cui, 2020).

3.2 Prediction layer

The prediction layer estimates a target such as course failure, dropout, low final performance, or early-risk status. Models should be evaluated not only by accuracy but also by recall for genuinely at-risk students, false-positive rates, precision, calibration, interpretability, and the earliest point at which performance is adequate for action (Cui, 2019; Howard et al., 2018).

3.3 Action layer

The action layer connects a prediction to a support pathway. A high-risk classification should trigger review rather than automatic labeling. Advisors or instructors can examine the features that led to the alert, combine them with contextual knowledge, and select an appropriate support response. This reduces the risk that a model becomes a self-fulfilling label.

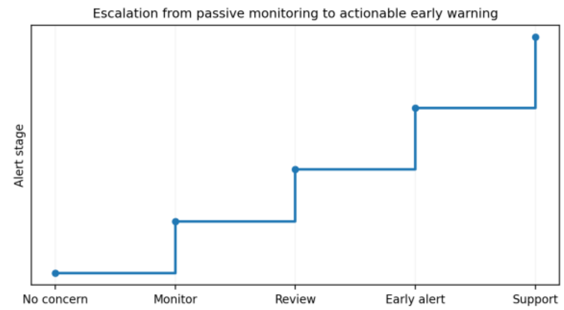

Figure 6. Escalation from passive monitoring to actionable early warning.

4. Methodology

4.1 Review design

This paper adopts a structured literature review and thematic evidence-synthesis design. The review focuses on studies that use learning analytics, educational data mining, predictive modeling, or early-warning approaches to identify academic risk or support student success. The evidence boundary ends at 2021, consistent with the study design, and the reviewed corpus contains 30 studies selected for direct relevance and open or openly accessible full text.

4.2 Search and selection strategy

Search concepts included combinations of 'learning analytics', 'early warning system', 'at-risk students', 'student success', 'dropout prediction', 'academic performance prediction', 'educational data mining', 'clickstream', 'LMS data', 'predictive analytics', and 'machine learning'. Searches were oriented toward peer-reviewed journal and conference research with a clear learning-analytics or predictive-education component. Studies published in 2022 or later were excluded from the synthesis.

4.3 Inclusion criteria

I. Published in 2021 or earlier.
II. Directly related to learning analytics, early warning, academic-risk prediction, dropout/retention prediction, or predictive student success.
III. Focused on higher education, online/distance education, or educational environments where the findings transfer clearly to academic-risk identification.
IV. Provided empirical, review, or methodological evidence relevant to prediction, early detection, intervention, or governance.
V. Accessible through an open-access publisher, repository, or openly available manuscript where possible.

a. Exclusion criteria

VI. Studies published from 2022 onward.

VII. Studies focused solely on non-educational predictive tasks.

VIII. Papers with no connection to student risk, performance, persistence, retention, or learning analytics.

IX. Duplicate records or papers that did not contain enough information to support thematic synthesis.

4.5 Coding framework

Table 2: Coding framework used for the 30-paper synthesis

Dimension	**Coding variables**	**Purpose**
Bibliographic	Year, venue, country/context	Descriptive profile
Data	LMS, assessment, administrative, interaction	Input signals
Target	Failure, dropout, retention, grade, risk	Outcome definition
Model	Regression, tree, kNN, ensemble, neural, time series	Prediction approach
Timing	Early weeks, mid-course, end-course	Action window
Intervention	Alert, advisor review, outreach, retention program	Actionability
Ethics	Privacy, transparency, fairness, governance	Responsible use

4.6 Analytical approach

The analysis used thematic synthesis rather than a pooled statistical meta-analysis because the reviewed studies differ substantially in outcome definitions, features, data structures, modeling methods, and performance metrics. Results were organized around data sources, prediction methods, timing, risk indicators, interventions, and governance. Figures in the Results section that contain numeric values are explicitly identified as author-coded thematic summaries or conceptual illustrations rather than pooled effect sizes.

5. Results

5.1 Corpus profile

The reviewed literature forms a field in which predictive modeling, early-warning systems, retention analysis, and learning analytics increasingly overlap. The corpus includes foundational proof-of-concept work, institutional systems, predictive-model comparisons, intervention evaluations, and recent systematic reviews. Macfadyen and Dawson (2010) provided an early LMS-based warning-system model, while later studies broadened the scope to online courses, distance education, administrative records, and machine-learning ensembles (Wolff et al., 2014; Berens et al., 2019; Karalar et al., 2021).

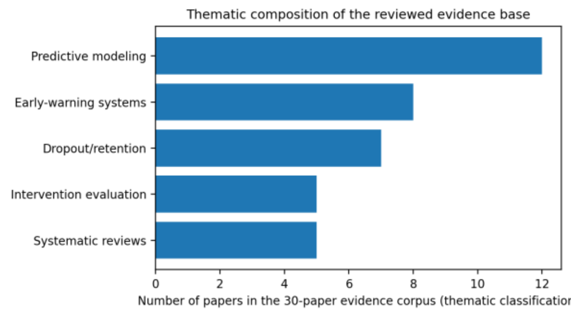

Figure 7. Thematic composition of the reviewed evidence base.

5.2 Recurring predictors of academic risk

Across the corpus, the most persistent predictors are related to engagement and demonstrated performance. Low or declining LMS activity, weaker assessment results, late or missing submissions, and irregular participation recur across different methodological designs. Casey and Azcona (2017) showed that student activity patterns can support performance prediction, while Aljohani et al. (2019) demonstrated the usefulness of clickstream data for identifying at-risk students in a virtual learning environment. These patterns reinforce the value of longitudinal behavior rather than one-time demographic or static variables alone.

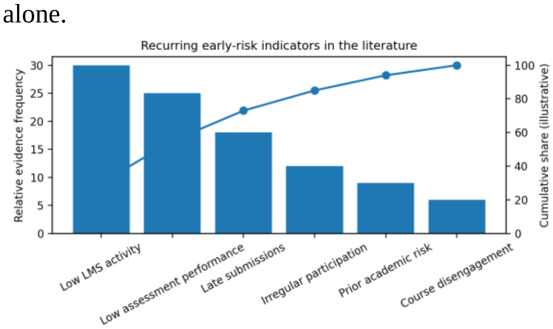

Figure 8. Recurring early-risk indicators reported across the literature.

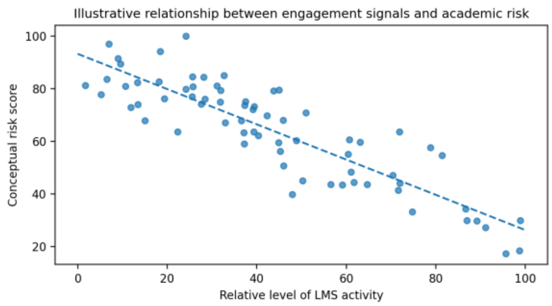

Figure 9. Illustrative relationship between learning activity and academic risk.

5.3 Data integration

A major pattern is the movement from single-source analysis toward integrated data. LMS logs are often highly available but can be ambiguous: low activity can indicate disengagement, but it can also reflect use of external materials or efficient study behavior. Assessment records provide outcome-relevant evidence, while administrative variables can add historical context. The combination of multiple sources can reduce overreliance on any single behavioral proxy (Ifenthaler & Yau, 2020; Albreiki et al., 2021).

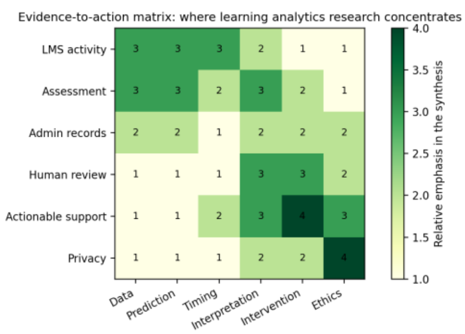

Figure 10. Evidence-to-action matrix showing where the reviewed literature places emphasis across the learning-analytics pipeline.

5.4 Prediction methods

The reviewed studies use a diverse modeling toolkit. Macfadyen and Dawson (2010) established a statistical early-warning approach; Howard et al. (2018) compared alternative prediction methods; Akçapınar et al. (2019) compared algorithms and preprocessing; and Karalar et al. (2021) used ensemble learning. Namoun and Alshanqiti (2021) systematically review machine-learning and data-mining methods and show that prediction performance is highly dependent on dataset characteristics, features, and evaluation choices. Consequently, a responsible methodological recommendation is not to declare one universally best algorithm but to select a model according to the target outcome, time horizon, sample size, interpretability requirements, and intervention context.

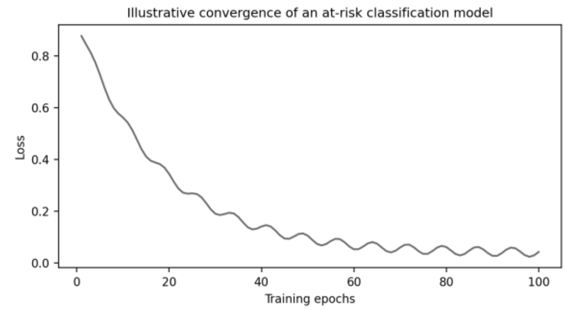

Figure 11. Illustrative convergence of an at-risk classification model.

5.5 Prediction timing

The evidence consistently makes timing consequential. Wolff et al. (2014) and Akçapınar et al. (2019) explicitly explore whether course outcomes can be predicted before the end of a learning period. Chen and Cui (2020) show the importance of time-series behavior, while Bañeres et al. (2019) examine early feedback prediction for students at risk. Taken together, these studies support defining an 'actionable prediction point' as the earliest moment at which model performance is adequate and sufficient time remains for educational support.

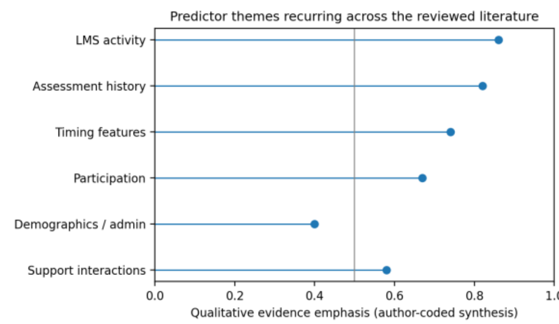

Figure 12. Recurring predictor themes across the reviewed literature.

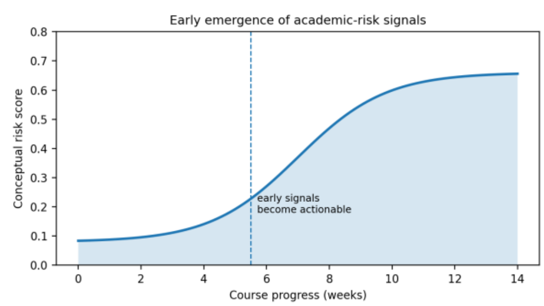

Figure 13. Conceptual early emergence of academic risk revisited as a result interpretation.

5.6 Outcome targets

Studies define success and risk through several targets: course failure, final performance, dropout or withdrawal, retention, and early-alert status. Ifenthaler and Yau (2020) emphasize the breadth of 'study success,' whereas de Oliveira et al. (2021) focus specifically on dropout prevention. This diversity means that early-warning systems should declare their target outcome explicitly. A model trained to predict final grade may not be suitable for dropout prevention, and a model trained on one course may not transfer to another without recalibration.

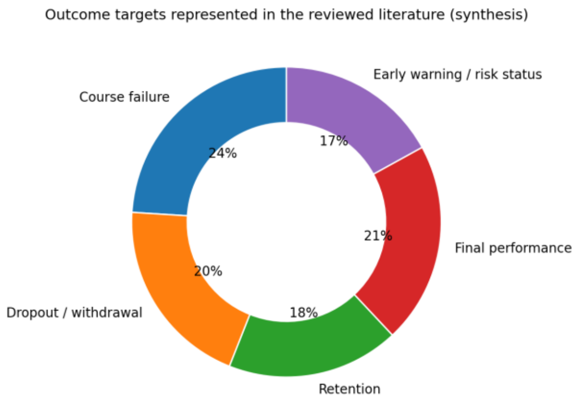

Figure 14. Outcome targets represented in the early-identification literature.

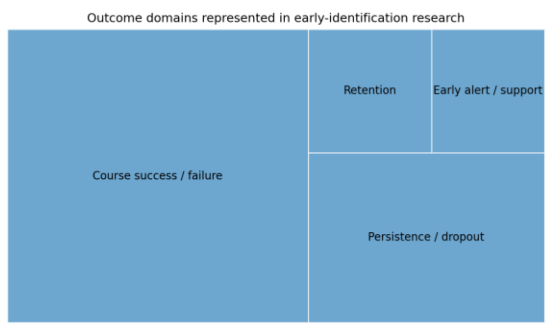

Figure 15. Outcome domains represented in early-identification research.

5.7 Prediction error and uncertainty

False positives and false negatives have different educational costs. A false negative may leave a genuinely struggling student without support, whereas a false positive may unnecessarily increase outreach or stigmatize a learner. Model evaluation should therefore include confusion-matrix measures, calibration, class imbalance handling, and the institutional cost of different types of prediction error. Namoun and Alshanqiti (2021) and Albreiki et al. (2021) both identify methodological inconsistency as a challenge in comparing predictive studies.

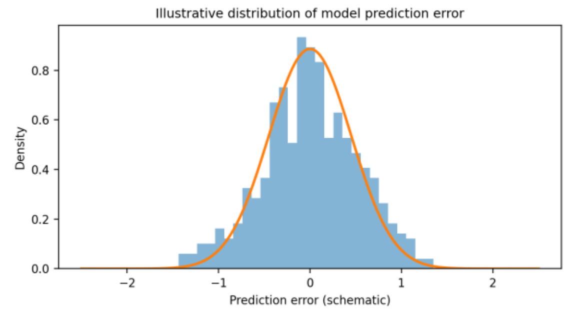

Figure 16. Illustrative prediction-error distribution.

5.8 From early warning to student support

The literature suggests a progression from prediction to graduated response. A useful system can begin with passive monitoring, move to advisor review when a pattern becomes concerning, generate an early alert when evidence is sufficiently strong, and trigger targeted academic support. This graduated structure reduces the risk of treating every deviation as a crisis and helps match support intensity to evidence strength.

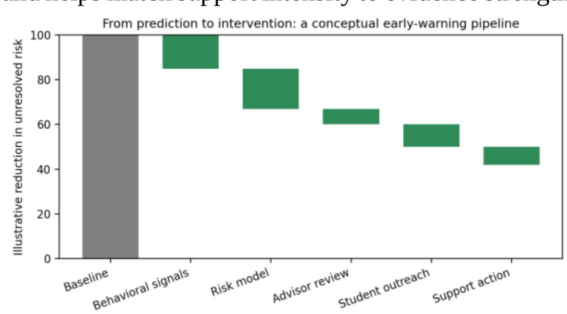

Figure 17. From prediction to intervention: conceptual reduction of unresolved academic risk through successive support stages.

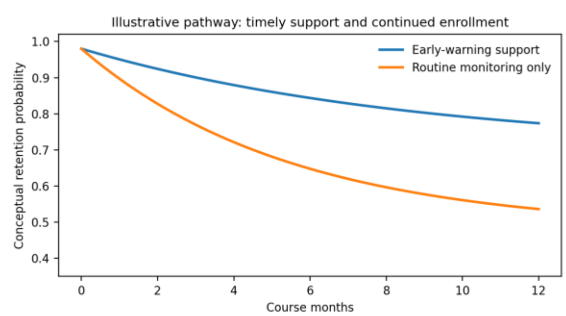

Figure 18. Illustrative retention pathway contrasting timely support with routine monitoring.

6. Discussion

6.1 Learning analytics works best as a decision-support system

The primary implication of the synthesis is that learning analytics should be designed as decision support rather than automated judgment. Predictive models identify patterns that deserve attention; they do not determine a student's future. This interpretation is consistent with the evolution from LMS data mining (Macfadyen & Dawson, 2010) to institutional early-alert systems (Jayaprakash et al., 2014) and intervention-oriented programs (Dawson et al., 2017). The educational objective is therefore to improve the speed and quality of human response, not to replace human judgment.

6.2 Multimodal data are more meaningful than isolated signals

The synthesis indicates that the most interpretable systems combine evidence. An abrupt decline in LMS activity becomes more informative when it co-occurs with falling assessment performance and missed submissions. Similarly, a student with low login frequency may not actually be at risk if assessment performance remains strong. This contextual logic cautions against simplistic thresholds based on a single activity measure. Multisource and longitudinal representations are more consistent with the complex nature of student success (Ifenthaler & Yau, 2020; Chen & Cui, 2020).

6.3 Early is not automatically better

The phrase 'early identification' can be misleading if early predictions are too uncertain to act upon. There is a balance between prediction lead time and predictive reliability. A signal generated during the first week may be weak because little behavioral history exists; a highly reliable prediction made immediately before final assessment may be too late. The appropriate goal is therefore the earliest actionable point, defined jointly by model performance and intervention feasibility.

6.4 Prediction should become intervention intelligence

A mature early-warning system should eventually move beyond estimating who is at risk to estimating which support action is appropriate and likely to help. Olaya et al. (2020) illustrate the value of uplift modeling for dropout prevention, which shifts attention toward differential intervention effects. This direction can reduce inefficient outreach and create a more personalized student-support process. Future systems should therefore capture intervention history and subsequent outcomes so that institutions can learn which actions work for which student contexts.

6.5 Explainability is educationally consequential

In a purely technical application, predictive accuracy may dominate model selection. In education, however, an opaque prediction can be difficult for instructors to trust and difficult for students to challenge. Explainability allows advisors to see whether a flag was driven by declining activity, missing submissions, weak assessment results, or another signal. It also enables institutions to test whether a model is behaving as intended. The practical value of a model is thus partially determined by the quality of the explanation accompanying the prediction.

6.6 Ethical use and student agency

Risk labeling can create unintended consequences if students are treated as categories rather than participants in the support process. Jones (2019) argues for consent mechanisms that protect student autonomy, while Ifenthaler et al. (2021) emphasize educationally meaningful and responsible use of analytics. A transparent system should communicate that a risk score is probabilistic, give students access to appropriate support, allow contest or correction of inaccurate data, and avoid using predictive risk as a punitive measure.

6.7 Proposed data-to-action model

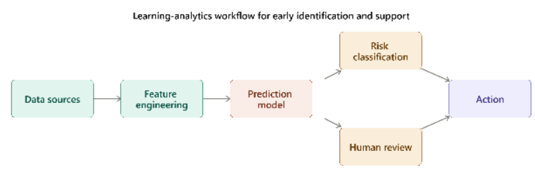

Figure 19. Integrated data-to-action model for early identification and academic support.

The proposed model has seven functional stages: (1) collect longitudinal educational traces; (2) engineer features that represent change and persistence; (3) generate risk estimates; (4) interpret the estimate with contextual evidence; (5) conduct human review; (6) provide targeted support; and (7) evaluate the outcome of both the student and the intervention. This final feedback stage is essential because it allows an institution to learn whether its early-warning process is actually improving student success. Without outcome feedback, the system can remain trapped in a loop of increasingly sophisticated prediction without evidence of educational impact.

7. Practical Implications

7.1 For institutions

Institutions should begin with a clearly defined student-success problem before selecting an algorithm. Data governance, data quality, and intervention capacity should be established alongside the predictive pipeline. An institution that lacks advisors or support services may gain little from highly accurate prediction because it cannot act on the alerts. Early-warning programs should therefore be funded and evaluated as complete support systems.

7.2 For instructors and advisors

Alerts should be interpreted as prompts for inquiry. Staff should review the evidence underlying a flag, consider contextual factors, contact the student respectfully, and document the support offered. Instructor-facing dashboards should favor concise explanations and trends over complex model outputs. Waddington et al. (2016) and Jayaprakash et al. (2014) illustrate the importance of translating data into usable educational information.

7.3 For data scientists

Model development should explicitly address temporal leakage, class imbalance, missing data, calibration, subgroup performance, and model portability. A model that performs well in one course or institution may degrade when the learning environment changes. Periodic validation and drift monitoring should therefore be included in operational systems.

7.4 For policy and governance

Policies should specify why student data are collected, who may access predictions, how long data are retained, what students are told, and how erroneous predictions can be corrected. Predictive analytics should not be used as a substitute for human support or as a basis for punitive treatment. Responsible governance is part of predictive validity because students and staff must trust the system enough to use it appropriately.

8. Limitations

This review has several limitations. First, the literature is heterogeneous, preventing direct aggregation of model-performance metrics into a single overall effect size. Second, the 30-paper corpus ends in 2021 by design, so the synthesis does not represent research published after that boundary. Third, different studies define risk, success, and prediction horizons differently. Fourth, some studies emphasize technical prediction and provide limited evidence about intervention outcomes. Finally, the thematic figures in this paper are designed to communicate patterns and mechanisms; they are not a substitute for a meta-analysis of common quantitative measures.

9. Conclusion

Learning analytics can strengthen academic-support systems by identifying changes in student behavior and performance before final outcomes are determined. The reviewed literature demonstrates the feasibility of predicting course failure, dropout, retention, and academic performance from LMS activity, assessment data, administrative records, and other educational traces. At the same time, the evidence shows why prediction should not be confused with success. Early-warning systems generate educational value only when predictions are timely, interpretable, responsibly governed, and connected to actions that students can actually use.

The central contribution of this paper is a data-to-action perspective: data are transformed into features, features into risk estimates, risk estimates into human-reviewed alerts, alerts into targeted support, and support into measurable outcomes that feed back into system improvement. Future learning-analytics research should therefore evaluate not only model accuracy but also lead time, explanation quality, intervention uptake, student experience, equity, and post-alert outcomes. The most useful question is no longer simply whether institutions can identify at-risk students early; it is whether they can use early identification to create fair, timely, and effective opportunities for students to succeed.

References

Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. “International Journal of Educational Technology in Higher Education, 16”, Article 40. https://doi.org/10.1186/s41239-019-0172-z

Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student' performance prediction using machine learning techniques. Education Sciences, 11(9), 552. https://doi.org/10.3390/educsci11090552

Aljohani, N. R., Fayoumi, A., & Hassan, S.-U. (2019). Predicting at-risk students using clickstream data in the virtual learning environment. Sustainability, 11(24), 7238. https://doi.org/10.3390/su11247238

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. “In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge” (pp. 267–270). Association for Computing Machinery. https://doi.org/10.1145/2330601.2330666

Bañeres, D., Rodríguez-González, M. E., & Serra, M. (2019). An early feedback prediction system for learners at-risk within a first-year higher education course. IEEE Transactions on Learning Technologies, 12(2), 249–263. https://doi.org/10.1109/TLT.2019.2912167

Bañeres, D., Rodríguez, M. E., Guerrero-Roldán, A. E., & Karadeniz, A. (2020). An early warning system to detect at-risk students in online higher education. Applied Sciences, 10(13), 4427. https://doi.org/10.3390/app10134427

Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2019). Early detection of students at risk—Predicting student dropouts using administrative student data from German universities and machine learning methods. Journal of Educational Data Mining, 11(3), 1–41. https://doi.org/10.5281/zenodo.3594771

Casey, K., & Azcona, D. (2017). Utilizing student activity patterns to predict performance. International Journal of Educational Technology in Higher Education, 14, Article 4. https://doi.org/10.1186/s41239-017-0044-3

Chen, F., & Cui, Y. (2020). Utilizing student time series behaviour in learning management systems for early prediction of course performance. Journal of Learning Analytics, 7(2), 1–17. https://doi.org/10.18608/jla.2020.72.1

Choi, S. P. M., Lam, S. S., Li, K. C., & Wong, B. T. M. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. Educational Technology & Society, 21(2), 273–290.

Cui, Y. (2019). Predictive analytic models of student success in higher education: A review of methodology. Information and Learning Sciences, 120(3–4), 208–227. https://doi.org/10.1108/ILS-10-2018-0104

Dawson, S., Jovanović, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In "Proceedings of the 7th International Conference on Learning Analytics and Knowledge" (pp. 474–478). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027405

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), 64. https://doi.org/10.3390/bdcc5040064

Howard, E., Meehan, M., & Parnell, A. (2018). Contrasting prediction methods for early warning systems at undergraduate level. The Internet and Higher Education, 37, 66–75. https://doi.org/10.1016/j.iheduc.2018.02.001

Ifenthaler, D., Gibson, D., Prasse, D., Shimada, A., & Yamada, M. (2021). Putting learning back into learning analytics: Actions for policy makers, researchers, and practitioners. Educational Technology Research and Development, 69, 2131–2150. https://doi.org/10.1007/s11423-020-09909-8

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. Educational Technology Research and Development, 68, 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. Journal of Learning Analytics, 1(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Jones, K. M. L. (2019). Learning analytics and higher education: A proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. International Journal of Educational Technology in Higher Education, 16, Article 24. https://doi.org/10.1186/s41239-019-0155-0

Karalar, H., Kapucu, C., & Gürüler, H. (2021). Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system. International Journal of Educational Technology in Higher Education, 18, Article 63. https://doi.org/10.1186/s41239-021-00300-y

Lu, O. H. T., Huang, A. Y. Q., & Yang, S. J. H. (2021). Impact of teachers' grading policy on the identification of at-risk students in learning analytics. "Computers & Education, 163", 104109. https://doi.org/10.1016/j.compedu.2020.104109

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an 'early warning system' for educators: A proof of concept. Computers & Education, 54(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. Computers & Education, 103, 1–15. https://doi.org/10.1016/j.compedu.2016.09.005

Musso, M. F., Rodríguez Hernández, C. F., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: A machine-learning approach. Higher Education, 80, 875–894. https://doi.org/10.1007/s10734-020-00520-7

Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237. https://doi.org/10.3390/app11010237

Olaya, D., Vasquez, J., Maldonado, S., Miranda, J., & Verbeke, W. (2020). Uplift modeling for preventing student dropout in higher education. Decision Support Systems, 134, 113320. https://doi.org/10.1016/j.dss.2020.113320

Queiroga, E. M., Lopes, J. L., Kappel, K., Aguiar, M., Araújo, R. M., Munoz, R., Villarroel, R., &

Cechinel, C. (2020). A learning analytics approach to identify students at risk of dropout: A case study with a technical distance education course. Applied Sciences, 10(11), 3998. https://doi.org/10.3390/app10113998

Russell, J.-E., Smith, A., & Larsen, R. (2020). Elements of success: Supporting at-risk student resilience through learning analytics. Computers & Education, 152, 103890. https://doi.org/10.1016/j.compedu.2020.103890

Waddington, R. J., Nam, S., Lonn, S., & Teasley, S. D. (2016). Improving early warning systems with categorized course resource usage. Journal of Learning Analytics, 3(3), 263–290. https://doi.org/10.18608/jla.2016.33.13

Wolff, A., Zdrahal, Z., Herrmannova, D., Kuzilek, J., & Hlosta, M. (2014). Developing predictive models for early detection of at-risk students on distance learning modules. In Proceedings of the Machine Learning and Learning Analytics Workshop at LAK14.

Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. The Internet and Higher Education, 27, 44–53. https://doi.org/10.1016/j.iheduc.2015.05.002