International Journal of Enhanced Research in Educational Development (IJERED)
ISSN: 2320-8708, Vol. 14 Issue 1, Jan.-Feb., 2026

Learning Analytics for Predicting Student Success in Postgraduate Programmes

Rizwan J. Ahmad

ABSTRACT

The expansion of digital learning environments has enabled higher education institutions to collect increasingly detailed information about students' academic histories, assessment performance, engagement patterns, learning behaviours, interactions, and progression. Learning analytics applies statistical, computational, and educational methods to these data to understand learning processes, predict academic outcomes, and support timely interventions. This review critically examines the use of learning analytics for predicting student success in postgraduate programmes. It focuses on the conceptualisation of postgraduate success, principal data sources, predictive variables, analytical techniques, early-warning systems, dashboards, personalised interventions, ethical issues, institutional readiness, and research limitations. The review follows a structured narrative approach and synthesises findings from systematic reviews, empirical studies, predictive-model evaluations, and postgraduate-specific research. Evidence suggests that prior academic performance, assessment results, regularity of engagement, interaction with learning management systems, research progress, self-regulation, psychosocial characteristics, and academic support can contribute to predicting postgraduate outcomes.

Machine-learning approaches, including logistic regression, decision trees, random forests, support vector machines, neural networks, gradient boosting, and ensemble methods, have demonstrated predictive potential. Nevertheless, high predictive accuracy does not automatically improve student learning, retention, dissertation completion, or degree attainment. Learning analytics becomes educationally valuable only when predictions are interpretable, delivered early enough, connected to appropriate human interventions, and evaluated against meaningful postgraduate outcomes. Concerns relating to privacy, consent, algorithmic bias, transparency, surveillance, data quality, and the labelling of students require particular attention. The review concludes that learning analytics should complement rather than replace academic judgement. Future research should develop postgraduate-specific, theory-informed, explainable, equitable, and longitudinal models that integrate coursework, research milestones, supervisory engagement, well-being, and professional learning while maintaining strong ethical governance.

Keywords: learning analytics, predictive analytics, postgraduate education, student success, machine learning, academic performance, student retention, educational data mining, early-warning systems.

INTRODUCTION

Postgraduate education plays a central role in producing researchers, academics, professionals, administrators, innovators, and highly specialised workers. Unlike many undergraduate programmes, postgraduate education frequently combines advanced coursework with independent research, professional practice, dissertations, theses, clinical responsibilities, laboratory activity, fieldwork, and sustained supervision. Postgraduate students may also be employed full-time, studying at a distance, returning to education after a long interruption, or managing significant family and professional responsibilities. As a result, success in postgraduate education is multidimensional and cannot be represented adequately by examination grades alone.

Universities have traditionally identified struggling students through low marks, missed deadlines, prolonged registration, irregular attendance, or concerns raised by academic supervisors. These indicators are often detected after difficulties have become established. Digital learning systems now make it possible to identify patterns earlier by analysing academic records, assessment submissions, learning management system activity, library use, participation, communication, research milestones, and other forms of educational data. Learning analytics has therefore emerged as a potential means of moving student support from a reactive model toward more timely and preventive action.

Learning analytics is generally concerned with the measurement, collection, analysis, and interpretation of data about learners and their educational contexts for the purpose of understanding and improving learning. It is related to educational data mining, academic analytics, institutional research, artificial intelligence in education, and predictive modelling. However, learning analytics is distinguished by its explicit concern with educational improvement rather than prediction for its own sake. A technically accurate model that identifies students at risk but does not result in

appropriate support has limited educational value. The purpose of analytics should therefore extend from describing and forecasting educational outcomes to informing responsible action.

The higher-education literature has expanded rapidly since the early 2010s. Learning analytics has been used to identify disengagement, predict course performance, estimate dropout risk, recommend resources, support feedback, personalise learning, evaluate course design, and inform institutional decision-making. Yet systematic reviews continue to identify limited large-scale causal evidence that analytics itself improves learning outcomes. Much of the literature concentrates on undergraduate courses, massive open online courses, or individual online modules. Direct evidence concerning master's, doctoral, professional, and research-based postgraduate programmes remains comparatively underdeveloped (Ifenthaler & Yau, 2020; Ncube & Ngulube, 2024a).

This review analyses the potential and limitations of learning analytics for predicting success in postgraduate programmes. It addresses five principal questions. First, how should postgraduate student success be defined? Second, which data and variables are most useful for prediction? Third, what analytical techniques have been employed? Fourth, under what conditions do predictions produce meaningful educational interventions? Finally, what ethical, methodological, and institutional issues must be resolved before predictive learning analytics can be used responsibly?

Review Approach

The article adopts a structured narrative-review approach. It draws on peer-reviewed systematic reviews, conceptual papers, empirical investigations, experimental studies, institutional implementations, and postgraduate-specific research published mainly between 2011 and 2025. The literature was examined under the themes of conceptual foundations, indicators of success, data sources, predictive methods, intervention design, effectiveness, ethics, privacy, fairness, institutional implementation, and future research.

A structured narrative approach is appropriate because predictive learning analytics studies differ substantially in their definitions, data sources, algorithms, educational settings, time frames, and outcome measures. Some predict success at the level of an individual assessment, whereas others predict course completion, annual progression, retention, final grades, or degree completion. The models also report different performance measures, including accuracy, precision, recall, sensitivity, specificity, F1 score, and area under the receiver operating characteristic curve. These differences restrict direct statistical comparison and make critical interpretation necessary (Alyahyan & Düştegör, 2020; García-García et al., 2025).

The review gives particular attention to postgraduate education but also draws from the wider higher-education literature where postgraduate-specific evidence is unavailable. Findings from undergraduate studies are not treated as directly transferable. Instead, they are used to identify principles, methods, and risks that require validation in postgraduate contexts.

Conceptualising Learning Analytics

Learning analytics operates across four overlapping levels: descriptive, diagnostic, predictive, and prescriptive analytics. Descriptive analytics explains what has happened, such as identifying declining participation or missed activities. Diagnostic analytics investigates why a pattern may have occurred. Predictive analytics estimates the probability of a future outcome, such as course failure, delayed completion, withdrawal, or high academic achievement. Prescriptive analytics recommends or initiates an action intended to improve the predicted outcome.

Predictive learning analytics typically combines historical student data with current learning behaviour. A model may learn from previous cohorts that particular combinations of assessment performance, delayed submissions, irregular online access, weak engagement, and limited interaction are associated with non-completion. It then estimates the risk level of current students displaying similar patterns. These estimates may be communicated through dashboards, risk categories, alerts, progress reports, or adviser notifications.

The educational value of such predictions depends on the relationship between data, theory, and intervention. A correlation between learning management system clicks and grades does not necessarily explain how learning occurs. Frequent platform activity may indicate engagement, confusion, inefficient navigation, or simply the requirements of the course design. Similarly, a student who downloads resources and studies offline may appear inactive even while making substantial progress. Learning analytics should therefore be grounded in theories of self-regulated learning, student engagement, motivation, feedback, social integration, adult learning, and academic persistence.

Gašević, Dawson, and Siemens argued that learning analytics must remain focused on learning rather than becoming an exercise in technical optimisation. Predictive variables should be interpreted in relation to the instructional environment, because the same digital behaviour can have different meanings in different programmes. More recent meta-review evidence similarly shows that learning analytics studies often pay insufficient attention to pedagogical aims, disciplinary contexts, theoretical foundations, and students' stages of learning (Gašević et al., 2015; García-García et al., 2025).

4. Defining Student Success in Postgraduate Programmes

A fundamental challenge is defining the outcome to be predicted. In undergraduate analytics, student success is frequently operationalised as a passing grade, course completion, first-year retention, or degree completion. These indicators remain relevant to postgraduate programmes, but they do not capture the complexity of advanced education.

Postgraduate success may include satisfactory coursework grades, progression from taught modules to research, approval of a research proposal, timely completion of ethical clearance, data collection, publication, professional competence, dissertation submission, thesis examination, degree completion, employability, and the development of an academic or professional identity. For doctoral candidates, success may involve persistence through uncertainty, productive supervision, participation in scholarly communities, research independence, publication activity, and completion within an expected period.

The choice of outcome affects every stage of predictive modelling. A model designed to predict course grades may use assessment marks and online engagement. A model predicting thesis completion may need supervision records, milestone achievement, research-methods competence, funding status, employment load, changes in registration, and the timing of proposal approval. A model predicting professional success may require workplace assessment, clinical performance, reflective portfolios, or competency data.

Success should not be reduced automatically to institutional efficiency. A student who completes quickly is not necessarily more successful than a student who takes additional time to produce rigorous research while managing employment or personal responsibilities. Similarly, delayed completion does not always represent failure. Postgraduate models must distinguish between temporary interruption, approved leave, part-time progression, administrative delay, research complexity, and genuine academic disengagement.

Alyahyan and Düştegör emphasised that predictive work should begin with a clear definition of academic success. Their review identified degree-level, year-level, course-level, and examination-level prediction as distinct analytical tasks. The lack of consistent outcome definitions contributes to variation in reported model performance and restricts comparisons across studies (Alyahyan & Düştegör, 2020).

Data Sources for Postgraduate Learning Analytics

5.1 Academic and Administrative Records

Academic records are among the most frequently used predictive data. They include previous qualifications, undergraduate grades, entrance-test results, postgraduate coursework marks, cumulative grade point averages, failed modules, credit completion, enrolment status, interruptions, and repeated registrations. Prior academic performance is consistently identified as one of the strongest predictors of later achievement.

For postgraduate students, undergraduate performance may indicate foundational disciplinary knowledge, academic literacy, study habits, and preparedness for advanced work. However, reliance on previous grades can reproduce historical disadvantage. Students from under-resourced institutions may have lower prior results despite possessing substantial potential. Prior attainment should therefore inform support rather than become an automatic basis for exclusion or reduced expectations.

A study of 1,091 postgraduate medical students used undergraduate academic information, psychosocial factors, and foundational skills to classify postgraduate performance. Undergraduate achievement emerged as the most influential predictor, while confidence, English proficiency, and other individual characteristics also contributed. The study combined logistic regression, XGBoost, feature selection, and Shapley Additive Explanations to improve interpretability (Lai et al., 2025).

5.2 Learning Management System Data

Learning management systems generate data concerning logins, page views, video access, downloads, quiz attempts, assignment submissions, forum participation, time stamps, activity sequences, and patterns of study. These data are attractive because they can be collected continuously and analysed before final assessments occur. Indicators such as regular access, timely completion of formative tasks, participation in assessment-related activities, and consistent study patterns may be more meaningful than total clicks. A student who accesses a course frequently but does not complete substantive tasks may not be learning effectively. Predictive features should therefore distinguish between administrative navigation, content engagement, assessment activity, and social interaction.

The Open University Learning Analytics Dataset has supported extensive research on the relationship between virtual learning environment activity, assessment, demographic characteristics, and course outcomes. Such datasets have helped researchers compare modelling techniques and examine temporal engagement. Nevertheless, models developed in large distance-learning environments may not transfer directly to small postgraduate cohorts, research degrees, laboratory programmes, or professional courses.

5.3 Assessment and Feedback Data

Assessment data include marks, submission timing, number of attempts, feedback use, rubric performance, plagiarism reports, quiz responses, and improvement across drafts. Early formative assessment is often a strong predictor because it reflects both prior knowledge and current engagement. In postgraduate education, assessment analytics can move beyond total scores. Rubric-level data may reveal difficulties in critical analysis, research design, theoretical integration, academic writing, evidence evaluation, quantitative reasoning, or professional decision-making. These insights can support targeted workshops or individual feedback. However, predictive use of assessment data should not encourage excessive standardisation of complex scholarly work.

5.4 Research Progress and Supervision Data

Research-based postgraduate programmes require additional data sources that are rarely represented in conventional analytics. Potential indicators include meeting frequency, milestone completion, proposal approval, ethics approval, chapter submissions, revision cycles, laboratory progress, fieldwork completion, publication activity, conference participation, supervisory feedback, and changes in the research plan.

These indicators require cautious interpretation. A low number of supervision meetings may reflect student independence, poor supervision, scheduling difficulties, or disciplinary norms. Frequent revisions may indicate weak performance or a productive process of intellectual development. Quantitative data should therefore be supplemented by contextual information and professional judgement.

5.5 Demographic, Psychosocial, and Contextual Data

Predictive models may include age, gender, employment, location, financial support, disability, language background, caregiving responsibilities, motivation, self-efficacy, confidence, belonging, stress, and study strategies. These variables can improve understanding of barriers but also create serious risks of discrimination and stereotyping.

Demographic variables often predict outcomes because they reflect structural inequalities rather than inherent ability. A model might associate part-time status or rural location with non-completion, but the appropriate response is not to label such students as deficient. Instead, the institution should examine whether scheduling, connectivity, supervision, financial support, or access to facilities is creating unequal conditions.

Reviews indicate that prior academic achievement and demographic information are among the most frequently used variables in higher-education prediction. However, their prominence does not mean that they are always the most educationally actionable variables. Data about current engagement, assessment progress, help-seeking, and supervisory contact may provide more useful opportunities for intervention (Alyahyan & Düştegör, 2020; García-García et al., 2025).

6. Analytical Techniques for Predicting Success

6.1 Statistical Models

Logistic regression is widely used to predict binary outcomes such as completion or withdrawal. Linear regression can estimate continuous outcomes such as grades or completion time. These methods are valued for their relative interpretability. Coefficients can show the direction and strength of relationships between variables and outcomes.

Statistical models may be suitable for small postgraduate cohorts where highly complex machine-learning approaches would overfit the data. They also facilitate explanation to academic staff. Their limitations include assumptions concerning linearity, independence, and variable relationships. Complex educational outcomes may involve interactions and nonlinear patterns that simpler models cannot capture.

6.2 Decision Trees and Random Forests

Decision trees divide students into groups based on combinations of predictive features. Their visual structure can be understandable to non-technical users. For example, a tree may indicate that students with low formative marks, late submissions, and irregular activity have a higher risk of non-completion.

Random forests combine multiple trees and frequently provide stronger predictive performance. They can model nonlinear relationships and interactions, but their collective decisions are less transparent than those of a single tree. Feature-importance scores can assist interpretation, although importance does not prove that a variable causes the predicted outcome.

6.3 Support Vector Machines and Neural Networks

Support vector machines can perform well when relationships between variables are complex and the number of predictors is large. Neural networks and deep-learning models can identify patterns in extensive datasets, including clickstreams, text, sequences, and multimodal data.

Deep-learning models have been used to predict academic performance from virtual learning environment data, but their complexity creates difficulties for interpretation. In postgraduate education, where decisions may affect supervision, progression, or continuation, an unexplained prediction is ethically and educationally problematic. A marginal improvement in accuracy may not justify a substantial loss of transparency.

6.4 Gradient Boosting and Ensemble Methods

Gradient-boosting approaches such as XGBoost build a series of models that correct previous errors. They frequently perform strongly with structured educational data. Ensemble methods combine predictions from multiple algorithms to improve robustness.

The use of explainable artificial intelligence has become increasingly important. Techniques such as SHAP can estimate how individual variables contribute to a particular prediction. The 2025 postgraduate medical study demonstrates how such methods can balance predictive performance with interpretability. Yet explainability tools should not be treated as complete explanations of learning. They explain the model's calculation, not necessarily the educational or causal process behind the outcome (Lai et al., 2025).

6.5 Natural Language and Social Network Analysis

Postgraduate education produces large amounts of textual and relational data through forum posts, reflective writing, dissertation drafts, feedback, emails, and supervision records. Natural language processing can identify themes, sentiment, linguistic complexity, argumentation, or patterns in feedback. Social network analysis can examine participation, collaboration, centrality, and isolation in learning communities.

These methods may be valuable for identifying students disconnected from academic networks or struggling with scholarly writing. However, written communication is sensitive and context-dependent. Automated sentiment analysis may misinterpret disciplinary critique, cultural communication styles, or second-language writing. Private supervisory communication should not be analysed without clear justification, consent, and governance.

7. Predictive Models, Dashboards, and Early-Warning Systems

Predictive results are commonly communicated through dashboards or early-warning systems. A student-facing dashboard may display progress, engagement, predicted grades, completed activities, or comparison with a cohort. A teacher-facing dashboard may identify students whose participation or assessment patterns differ from expected behaviour. An institutional system may estimate retention and resource needs.

Dashboards can support self-regulation by making progress visible. They may encourage students to review their behaviour, manage time, seek help, and complete missed tasks. Yet dashboards can also create anxiety, unproductive competition, or false confidence. A student classified as "low risk" may reduce effort, while a student shown a high probability of failure may disengage.

A randomised controlled experiment involving 556 students found that access to a learning analytics dashboard and weekly email positively affected online behaviour but did not improve final examination performance. The finding demonstrates that behavioural change does not necessarily translate into improved achievement. It also reinforces the need to distinguish between predictive accuracy, dashboard use, behavioural response, and educational outcome (Hellings & Haelermans, 2022).

Teacher engagement with analytics is equally important. A large-scale implementation involving 59 teachers, nine distance-learning courses, and 1,325 students found that teachers could positively influence performance when they engaged with predictive information and used it to support students. Interviews indicated that predictive data affected teaching practices and intervention strategies. The study suggests that analytics is most useful when embedded in relationships and professional action rather than delivered as an automated classification (Herodotou et al., 2019).

For postgraduate programmes, early-warning systems should be adapted to programme structure. A taught master's programme may require weekly or module-level monitoring. A doctoral programme may require milestone-based monitoring over several years. Continuous weekly activity may be less relevant during fieldwork or independent writing, while missed supervision, delayed ethics approval, repeated failure to meet agreed milestones, or long periods without research progress may be more meaningful.

8. From Prediction to Educational Intervention

The central question is not whether student outcomes can be predicted but whether prediction enables effective support. An intervention may involve a personalised message, adviser contact, academic-skills support, supervisory review, counselling, financial guidance, peer mentoring, writing assistance, methodological training, or adjustment to the study plan.

Interventions should be proportionate to the evidence. A small reduction in platform activity should not trigger an intrusive institutional response. Risk estimates should be combined with other information and, wherever possible, discussed with the student. The student may provide an explanation that is not visible in the data, such as temporary employment pressure, illness, fieldwork activity, offline study, or technical difficulties.

Timing is critical. Predictions must be produced early enough for intervention. A highly accurate model that identifies likely failure after the final assessment has little preventive value. Early models may be less accurate because limited current-course data are available. Institutions must therefore balance timeliness with reliability.

The intervention must also address a modifiable problem. Prior undergraduate grades may predict postgraduate outcomes but cannot be changed. In contrast, weak academic writing, irregular formative assessment, limited contact with a supervisor, or uncertainty about research methods may be addressed through support. Actionability should therefore be considered when selecting predictive variables.

Systematic reviews indicate that learning analytics offers promising techniques for identifying students at risk, but rigorous large-scale evidence concerning its effect on continuation and completion remains limited. The literature contains many predictive models but fewer controlled evaluations of the actions that follow prediction. In one major review, none of the included studies was rated methodologically strong, highlighting persistent weaknesses in theoretical clarity, research design, and evaluation (Ifenthaler & Yau, 2020).

Similarly, reviews of learning analytics interventions have reported generally promising but inconsistent outcomes. Variation occurs because interventions differ in intensity, timing, target population, communication style, staff involvement, and available support services. An alert alone is unlikely to resolve complex academic or personal difficulties. Effective intervention requires institutional capacity and a clearly defined pathway from identification to assistance.

9. Postgraduate-Specific Applications

The limited postgraduate literature suggests that data analytics may support progress monitoring, personalised learning, academic advising, research supervision, curriculum evaluation, and the identification of students requiring additional support. Ncube and Ngulube's systematic review examined the use of data analytics to enhance postgraduate academic achievement. It highlighted the potential of descriptive, predictive, and prescriptive analytics while recommending comprehensive evaluation systems, greater educator capacity, transparent communication, and stronger student data literacy (Ncube & Ngulube, 2024a).

Postgraduate prediction should recognise distinctions among programme types. In professional programmes, success may depend on workplace performance, competency assessment, clinical participation, and integration of theory with practice. In taught master's programmes, coursework engagement and assessment may provide useful early indicators. In research programmes, supervision, milestone progression, research environment, funding, and scholarly belonging may be more influential.

Small cohort sizes create a methodological difficulty. Machine-learning models often require large datasets, while specialist postgraduate programmes may admit only a small number of students each year. Combining several cohorts can increase sample size but may introduce changes in curriculum, assessment, staff, technology, and admission criteria. Combining different programmes can create heterogeneity that weakens interpretation.

Transferability is another problem. A model developed for an online business programme may perform poorly in a laboratory-based science programme. Even apparently similar courses may differ in teaching design. Research has shown that predictive relationships based on online trace data may have limited portability across learning environments and that students' internal conditions may explain outcomes beyond observable digital behaviour. Models should therefore be validated locally and monitored continuously rather than assumed to be universally applicable.

10. Ethical, Privacy, and Fairness Considerations

Predictive analytics involves collecting and interpreting information that can influence how students are perceived and treated. Ethical questions include what data should be collected, whether students have meaningfully consented, who can access predictions, how long information is retained, whether students can challenge conclusions, and what obligations arise when risk is identified.

Privacy is particularly important in postgraduate programmes because the data may include supervision records, research delays, psychological characteristics, financial information, disability, professional performance, workplace data, and sensitive communication. The fact that an institution can collect data does not mean that every available variable should be used.

Postgraduate-specific privacy research has emphasised the need to balance data utility with student rights through clear ethical frameworks, responsible governance, transparency, and privacy-preserving techniques. Institutions should identify the minimum data necessary for an educational purpose and avoid indefinite or unrelated secondary use (Ncube & Ngulube, 2024b).

Algorithmic bias is a further concern. Models learn from historical data, which may contain the effects of unequal access, biased assessment, inconsistent supervision, or structural discrimination. If historical patterns are reproduced uncritically, analytics may classify already disadvantaged groups as inherently risky. Predictions could then influence expectations, support, admission, funding, or progression in ways that reinforce inequality.

Fairness cannot be assessed using overall accuracy alone. Institutions should examine false positives and false negatives across different groups. A false positive occurs when a student is labelled at risk despite being likely to succeed. A false negative occurs when a student requiring support is not identified. Both have consequences. Unnecessary intervention can stigmatise or inconvenience students, while missed identification may deny needed support.

Transparency is necessary but technically difficult. Students should be informed about the purpose of analytics, data sources, model limitations, access arrangements, intervention process, and opportunities for correction or appeal. Academic staff should understand that predictions are probabilistic rather than definitive.

The DELICATE framework identifies institutional responsibility, purpose, legitimacy, stakeholder involvement, consent, anonymity, technical security, external partnerships, and communication as important dimensions of trusted learning analytics. Ethical research also emphasises care, justice, accountability, and the institutional obligation to act responsibly when analytics reveals potential difficulty (Drachsler & Greller, 2016; Slade & Prinsloo, 2013).

A systematic review of empirical ethics research found that the evidence base remained relatively limited, with only 21 studies meeting its criteria for the period examined. This indicates that ethical debate has developed more rapidly than empirical understanding of how students and teachers experience analytics in practice (Cerratto Pargman & McGrath, 2021).

11. Institutional Requirements for Responsible Implementation

Successful implementation requires collaboration among postgraduate schools, faculties, supervisors, students, data scientists, educational researchers, information-technology teams, legal specialists, ethics committees, and student-support services. Learning analytics cannot be treated solely as a software-purchasing decision.

Institutions should begin by defining the educational problem. A university concerned about doctoral non-completion should determine whether delays are related to admissions, supervision, funding, research design, administrative procedures, student well-being, or programme structure. Analytics should then be designed around plausible support mechanisms rather than around whatever data happen to be available.

Data quality is essential. Incomplete records, inconsistent coding, platform changes, missing offline activity, and differences among programmes can distort predictions. Data should be documented, cleaned, validated, and interpreted in relation to curriculum and policy.

Staff data literacy is equally important. Teachers and supervisors must understand prediction probabilities, uncertainty, model limitations, and the difference between correlation and causation. They should know how to communicate with students without presenting a risk score as a judgement of ability.

Students should be involved in design and evaluation. Their involvement can reveal whether indicators are understandable, whether interventions feel supportive or intrusive, and whether important aspects of learning are missing. Participatory design may also increase trust.

Finally, implementation should be evaluated as a complete system. Institutions should assess predictive performance, staff use, student response, intervention delivery, academic outcomes, equity, cost, and unintended consequences. A technically accurate model may fail because advisers do not use it, students distrust the messages, or suitable support is unavailable.

12. Limitations of the Existing Evidence

The learning analytics literature contains several methodological limitations. Many studies are retrospective and evaluate how accurately historical outcomes can be predicted. Retrospective accuracy does not show that the model will perform equally well with future cohorts or that interventions will improve results.

Sample size and class imbalance are persistent problems. In programmes with high completion rates, the number of unsuccessful students may be too small to train a reliable model. A model may achieve high overall accuracy simply by predicting that most students will succeed while failing to identify the minority at risk.

Data leakage can artificially inflate performance when variables closely connected to the final outcome are included in training. For example, a model that uses late-stage assessment marks to predict final grades may be accurate but provide little useful advance warning.

The selection of performance metrics can also create misleading conclusions. Accuracy alone is inadequate when classes are imbalanced. Researchers should report precision, recall, specificity, sensitivity, F1 scores, calibration, and performance across relevant student groups.

Theoretical weaknesses remain prominent. Models are frequently selected according to computational performance without a clear explanation of how variables represent learning, motivation, interaction, or educational context. The 2025 meta-review of 19 systematic reviews found critically low methodological quality and limited consideration of pedagogical aims, disciplinary context, theory, and student learning stages. The review concluded that heterogeneous definitions and metrics prevented robust general conclusions (García-García et al., 2025).

Postgraduate-specific evidence remains especially limited. Existing studies are concentrated in particular disciplines, institutions, or countries. Research-based master's and doctoral programmes are underrepresented, and few studies examine completion over several years. Consequently, institutions should avoid assuming that general undergraduate prediction models will be effective in postgraduate settings.

13. Future Research Directions

Future studies should develop multidimensional definitions of postgraduate success. Models should include academic achievement, research milestones, degree completion, professional competence, well-being, scholarly engagement, and equitable participation rather than relying exclusively on grades.

Longitudinal research is essential. Postgraduate journeys may extend over several years, and risk factors change over time. Early transition, coursework, proposal development, data collection, writing, examination, and completion represent different phases requiring different indicators and interventions.

Research should compare interpretable statistical models with more complex machine-learning methods. The objective should not be maximum accuracy alone. Model calibration, fairness, actionability, portability, and comprehensibility should be treated as central outcomes.

Explainable artificial intelligence requires further educational evaluation. SHAP and related tools can indicate how features influence model outputs, but researchers must determine whether these explanations help students, supervisors, and advisers make better decisions. Technical explanation is useful only when it leads to responsible understanding.

Intervention research should become a priority. Randomised, quasi-experimental, and mixed-method studies should examine which messages, dashboards, advising strategies, mentoring programmes, and supervisory interventions improve outcomes. Prediction and intervention must be evaluated separately.

Researchers should also investigate the emotional effects of predictive information. Risk messages may motivate some students while increasing anxiety or fatalism among others. Communication should be supportive, specific, and linked to achievable actions.

Cross-institutional validation is needed, but transfer should be handled carefully. Federated learning and privacy-preserving analytics may allow institutions to collaborate without transferring identifiable student records. These approaches could be valuable for small postgraduate cohorts, provided they do not obscure governance or accountability.

Finally, students should be recognised as partners rather than data subjects alone. Participatory governance, understandable consent, access to personal data, correction mechanisms, and the ability to contest predictions should become standard components of responsible postgraduate analytics.

CONCLUSION

Learning analytics has considerable potential to support the prediction and improvement of student success in postgraduate programmes. Academic history, assessment performance, engagement regularity, learning management system activity, research milestones, supervisory interaction, self-regulation, and psychosocial factors can contribute to identifying students who may require additional support. Statistical methods and machine-learning techniques can

generate useful predictive information, while dashboards and early-warning systems can communicate progress to students, teachers, supervisors, and advisers.

Nevertheless, prediction is not equivalent to improvement. Evidence shows that dashboards may alter online behaviour without improving academic performance, whereas teacher engagement and well-designed intervention may strengthen outcomes. Learning analytics should therefore be judged by whether it facilitates appropriate, timely, equitable, and effective educational action.

Postgraduate education requires models sensitive to disciplinary differences, research pathways, part-time study, professional responsibilities, supervision, and long-term progression. Models developed for undergraduate courses cannot simply be transferred to master's and doctoral programmes. High predictive accuracy is insufficient when the outcome is poorly defined, the data are biased, the model is opaque, or the institution lacks the capacity to intervene.

Responsible learning analytics should be human-centred, theory-informed, transparent, explainable, privacy-conscious, and subject to continuous evaluation. It should complement academic judgement rather than replace it. When implemented within strong ethical governance and connected to meaningful support, predictive learning analytics can help postgraduate institutions identify emerging difficulties, allocate resources more effectively, personalise academic assistance, and create conditions in which a wider range of students can succeed.

REFERENCES

1. Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education, 17*, Article 3. doi: 10.1186/s41239-020-0177-7
2. Banihashem, S. K., Noroozi, O., Van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review, 37*, 100489. doi: 10.1016/j.edurev.2022.100489
3. Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics, 6*(2), 1–9. doi: 10.18608/jla.2019.62.1
4. 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. doi: 10.18608/jla.2021.1
5. Drachsler, H., & Greller, W. (2016). Privacy and analytics—It is a DELICATE issue: A checklist for trusted learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 89–98). Association for Computing Machinery. doi: 10.1145/2883851.2883893
6. Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5–6), 304–317. doi: 10.1504/IJTEL.2012.051816
7. Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (pp. 56–65). Association for Computing Machinery. doi: 10.1145/3027385.3027396
8. García-García, F. J., Gómez-Núñez, M. I., & Molla-Esparza, C. (2025). Applications of learning analytics in the study of academic performance in higher education: A meta-review with an educational perspective. *Higher Education*. Advance online publication. doi: 10.1007/s10734-026-01709-y
9. Gašević, D., Dawson, S., & Siemens, G. (2015). Let us not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. doi: 10.1007/s11528-014-0822-x
10. Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one-size-fits-all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68–84. doi: 10.1016/j.iheduc.2015.10.002
11. Hellings, J., & Haelermans, C. (2022). The effect of providing learning analytics on student behaviour and performance in programming: A randomised controlled experiment. *Higher Education, 83*, 1–18. doi: 10.1007/s10734-020-00560-z
12. Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, 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*, 1273–1306. doi: 10.1007/s11423-019-09685-0
13. 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. doi: 10.1007/s11423-020-09788-z
14. Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education, 172*, 104251. doi: 10.1016/j.compedu.2021.104251
15. Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics Dataset. *Scientific Data, 4*, 170171. doi: 10.1038/sdata.2017.171

16. Lai, W., Lin, J., Shen, K., Ling, Z., & Guan, Y. (2025). Enhancing graduate education assessment: A machine learning-based classification of academic performance in medical students. *BMC Medical Education, 26*, Article 423. doi: 10.1186/s12909-026-08741-7
17. Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology, 50*(5), 2594–2618. doi: 10.1111/bjet.12720
18. Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 31–40.
19. Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies, 13*(2), 226–245. doi: 10.1109/TLT.2019.2916802
20. Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences, 11*(1), 237. doi: 10.3390/app11010237
21. Ncube, M. M., & Ngulube, P. (2024a). Optimising data analytics to enhance postgraduate student academic achievement: A systematic review. *Education Sciences, 14*(11), 1263. doi: 10.3390/educsci14111263
22. Ncube, M. M., & Ngulube, P. (2024b). A systematic review of postgraduate programmes concerning ethical imperatives of data privacy in sustainable educational data analytics. *Sustainability, 16*(15), 6377. doi: 10.3390/su16156377
23. Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49–64.
24. Pardo, A., Jovanović, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology, 50*(1), 128–138. doi: 10.1111/bjet.12592
25. Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analysing and predicting students' performance by means of machine learning: A review. *Applied Sciences, 10*(3), 1042. doi: 10.3390/app10031042
26. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery, 10*(3), e1355. doi: 10.1002/widm.1355
27. Sclater, N. (2016). Developing a code of practice for learning analytics. *Journal of Learning Analytics, 3*(1), 16–42. doi: 10.18608/jla.2016.31.3
28. Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 30–40.
29. Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. doi: 10.1177/0002764213479366
30. Tsai, Y. S., Poquet, O., Gašević, D., Dawson, S., & Pardo, A. (2019). Complexity leadership in learning analytics: Drivers, challenges and opportunities. *British Journal of Educational Technology, 50*(6), 2839–2854. doi: 10.1111/bjet.12846
31. Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98–110. doi: 10.1016/j.chb.2018.07.027
32. Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from virtual learning environment big data using a deep-learning model. *Computers in Human Behavior, 104*, 106189. doi: 10.1016/j.chb.2019.106189