Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Learning Analytics and Early Warning Systems for Mathematics Achievement: A Multi-Country Comparative Programme Evaluation

Maryam Jimoh

jimohmaryam530@gmail.com

Ladoke Akintola University of Technology https://orcid.org/0009-0005-4946-5890

Research Article

Keywords: learning analytics, early warning systems, mathematics achievement, predictive analytics, comparative education, algorithmic fairness, programme evaluation

Posted Date: August 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10559564/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Mathematics underachievement remains one of the most persistent challenges facing basic and higher education systems worldwide, and it is increasingly addressed through learning analytics (LA) and early warning systems (EWS) that convert routinely collected academic data into actionable risk indicators. This review synthesises evidence from thirty peer-reviewed and institutional studies published mainly between 2019 and 2026 to evaluate how LA-EWS programmes have been designed, implemented and assessed for mathematics achievement across diverse national contexts, including Chile, Germany, the United Kingdom, Estonia, Mexico, Türkiye and several cross-national assessment regimes. The review adopts a structured narrative-synthesis methodology informed by systematic review conventions, comparing predictor variables, modelling approaches, reported discriminative performance and equity outcomes. Findings indicate that discrimination, expressed as the area under the receiver operating characteristic curve, typically ranges between 0.70 and 0.93, yet operational usefulness depends heavily on prevalence stability, threshold governance and the extent to which indicators capture domain-specific mathematical misconceptions rather than generic engagement signals. Persistent gaps are identified in algorithmic fairness, cross-country comparability, and the integration of conceptual diagnostic evidence, particularly around student difficulties with functions and graphical representation. The review further considers how emerging acceptance of generative artificial intelligence tools among tertiary populations is likely to reshape future analytics adoption. The paper concludes with a comparative framework and policy-relevant recommendations for designing mathematics-specific, equity-aware early warning programmes suitable for cross-country evaluation.

1. Introduction

Mathematics performance continues to function as a gatekeeper for participation in higher education and technical occupations, yet international assessment cycles repeatedly document wide and persistent achievement gaps within and between education systems (Mullis et al., 2020; OECD, 2023). Digitalisation of teaching and administrative infrastructure has, over the past decade, generated an unprecedented volume of trace data describing how students interact with learning management systems, complete formative assessments and progress through curricular sequences. Learning analytics (LA) emerged as the disciplinary response to this data abundance, defined broadly as the measurement, collection, analysis and reporting of data about learners and their contexts for the purposes of understanding and optimising learning (Siemens, 2013). Within this broader field, early warning systems (EWS) constitute a specific and increasingly consequential application: predictive models that flag students at elevated risk of underperformance or withdrawal sufficiently early for pedagogical intervention to remain feasible.

Mathematics occupies a distinctive position within LA-EWS scholarship. Unlike broadly defined course-completion or retention outcomes, mathematics achievement is shaped by well-documented conceptual progressions and recurring misconceptions, particularly around functions, graphical reasoning and algebraic manipulation (Ayeh, 2025). This domain specificity means that generic engagement-based

predictors, however statistically powerful, may fail to capture the substantive reasons a student is struggling, limiting the diagnostic and pedagogical value of an alert even when its statistical discrimination is high. At the same time, the institutional contexts in which LA-EWS programmes are deployed vary substantially across countries in terms of data infrastructure, governance capacity, assessment design and cultural attitudes toward algorithmic decision-making in education (Hilliger et al., 2020). A comparative evaluation lens is therefore necessary to understand which design features of LA-EWS programmes generalise across settings and which remain contingent on local conditions.

This review addresses three interrelated aims. First, it synthesises the conceptual and methodological foundations of LA and EWS scholarship as they apply to mathematics achievement. Second, it constructs a comparative evidence base drawn from multi-country programme evaluations, examining predictor design, modelling choices and reported effectiveness. Third, it critically appraises the equity, governance and stakeholder-acceptance dimensions that determine whether technically sound early warning models translate into improved mathematics outcomes in practice. In pursuing these aims, the review also considers an adjacent and rapidly evolving development: the growing use of generative artificial intelligence tools by students and staff, and the implications of variable awareness and acceptance of such tools for the broader analytics ecosystem within which EWS operate (Ayeh and Kumah, 2026).

The remainder of the paper is organised as follows. Section 2 details the review methodology. Section 3 presents the literature review across seven thematic strands. Section 4 articulates the problem statement motivating the comparative evaluation. Section 5 reports the descriptive and comparative results of the synthesis, supported by three original tables and figures. Section 6 discusses the findings in relation to theory, policy and mathematics-specific pedagogy. Section 7 outlines the principal challenges and limitations of both the reviewed evidence base and the review itself. Section 8 concludes with recommendations for researchers, developers and policymakers.

2. Methodology

This review employs a structured narrative synthesis informed by systematic review conventions, appropriate for a comparative evaluation spanning heterogeneous national contexts, outcome definitions and reporting formats that preclude formal meta-analysis. The approach follows a transparent, replicable search-and-screening protocol while permitting thematic and comparative synthesis rather than pooled effect estimation.

2.1 Search Strategy and Sources

Electronic searches were conducted across Scopus, Web of Science, IEEE Xplore, ERIC and Google Scholar for records published between January 2015 and July 2026, with priority given to publications from 2019 onward. Search strings combined descriptors for learning analytics, predictive analytics, early warning systems and academic risk with mathematics-specific and comparative-education qualifiers

(for example, "learning analytics" AND "early warning" AND "mathematics"; "predictive model" AND "at-risk students" AND "STEM"; "cross-national" AND "academic achievement" AND "dashboard"). Reference lists of retrieved systematic reviews were hand-searched to identify additional eligible studies, and institutional and government reports (for example, OECD and TIMSS technical reports) were included where they provided cross-country comparative data relevant to mathematics achievement.

2.2 Eligibility Criteria

Studies were eligible for inclusion if they: (a) reported an empirical evaluation, design description or systematic synthesis of a learning analytics or predictive analytics application in an educational setting; (b) addressed mathematics achievement directly or reported findings generalisable to quantitative or STEM disciplines with clearly disaggregated mathematics-relevant data; (c) were published in English in a peer-reviewed journal, indexed conference proceeding or authoritative institutional report; and (d) provided sufficient methodological detail to characterise predictor variables, modelling approach and outcome metrics. Studies confined to non-educational domains, purely theoretical papers without an applied analytics component, and sources lacking verifiable authorship or venue were excluded.

2.3 Screening, Selection and Synthesis

Titles and abstracts were screened against the eligibility criteria, followed by full-text assessment of shortlisted records. The final corpus comprised thirty studies and reports spanning institutional case studies, systematic reviews, and cross-national assessment programmes across more than ten education systems, including Chile, Germany, the United Kingdom, Estonia, Mexico, Türkiye and OECD- and IEA-coordinated multi-country assessments. Data extracted from each source included country or regional context, education level, data sources and predictor categories, modelling technique, reported discriminative or effect-size metrics, and equity-relevant findings. Extracted data were organised into a comparative matrix (summarised in Tables 1 and 2) and synthesised thematically to identify recurring patterns, contradictions and gaps across the seven themes reported in the literature review.

2.4 Analytical Approach to Comparative Evaluation

Because the reviewed studies employ heterogeneous outcome definitions (course pass/fail, dropout, standardised test performance, formative assessment scores) and reporting conventions, comparability was achieved through narrative and visual synthesis rather than statistical pooling. Reported discrimination statistics, where available, were compared descriptively across country contexts (Fig. 2), and predictor categories were coded into seven mutually exclusive groups to characterise prevalence across the corpus (Fig. 3). This approach follows established practice in comparative educational technology reviews, which favour structured qualitative synthesis over meta-analysis when included studies differ substantially in design and metric (Ifenthaler and Yau, 2020).

3. Literature Review

3.1 Conceptual Foundations of Learning Analytics

Learning analytics was formally established as a distinct field with Siemens' (2013) characterisation of the discipline as the measurement and analysis of learner data to understand and optimise learning environments. Gašević, Dawson and Siemens (2015) subsequently cautioned that technical sophistication should not eclipse pedagogical grounding, arguing that analytics interventions are only as effective as the learning theory underpinning their design. This admonition remains salient: a recurring critique across the corpus reviewed here is that many LA-EWS implementations privilege readily available behavioural traces, such as login frequency, over theoretically motivated indicators of conceptual understanding. Viberg et al. (2018) traced the diffusion of LA across European higher education and observed that adoption has outpaced the accumulation of robust evidence regarding pedagogical impact, a tension that persists in the mathematics-specific literature examined below. An early institutional review of United Kingdom and international practice reached a similar conclusion at the sector level, finding that investment in analytics infrastructure had frequently outpaced the development of clear institutional strategies for acting on the resulting insights (Sclater, Peasgood and Mullan, 2016).

3.2 Origins and Evolution of Early Warning Systems

Institutional early warning practice was substantially advanced by Course Signals at Purdue University, where Arnold and Pistilli (2012) reported reductions in D, F and withdrawal rates following the combination of predictive scoring with timely instructor and student notification. The Open Academic Analytics Initiative extended this model across institutions, with Jayaprakash et al. (2014) demonstrating that models trained at one institution could transfer to others with acceptable, though diminished, predictive performance. Within mathematics-adjacent STEM courses specifically, Marbouti, Diefes-Dux and Madhavan (2016) compared seven prediction techniques using within-semester assessment data and found that partial-course grades reliably outperformed demographic variables as predictors of final outcomes, a finding echoed by later studies in this review. Akçapınar, Altun and Aşkar (2019) further demonstrated that automated notification mechanisms tied to real-time grade thresholds meaningfully support timely academic intervention, reinforcing the field's consensus that actionable timing is as important as predictive accuracy. Comparative modelling work has also examined the trade-off between predictive accuracy and interpretability: Beaulac and Rosenthal (2019) found that random forest classifiers modestly outperform logistic regression in predicting undergraduate students' academic success and choice of major, yet at the cost of the transparency that many institutions value when communicating risk classifications to students and staff.

3.3 Dashboards, Self-Regulated Learning and Stakeholder Use

Institutional early warning practice operationalises a much older theoretical proposition. Tinto (1975) argued that student persistence is shaped by the degree of a student's academic and social integration within an institution, a proposition that provides the conceptual justification for treating observable performance signals, such as partial-course grades, as proxies for underlying integration and risk of

disengagement. This theoretical anchor recurs implicitly throughout the applied EWS literature reviewed below, even where it is not always explicitly cited.

A substantial body of work examines how LA dashboards are used once risk information is generated. Matcha, Gašević and Pardo (2019) conducted a systematic review of dashboard studies from a self-regulated learning perspective, concluding that most dashboards support monitoring but rarely scaffold the planning and evaluation phases of self-regulation. Chen et al. (2023) found that high-performing students used dashboards more frequently during preview and review phases than lower-performing peers, suggesting that dashboards may inadvertently widen rather than narrow achievement gaps unless deliberately designed for at-risk users. Susnjak, Ramaswami and Mathrani (2022) proposed actionable-insight dashboards that move beyond descriptive visualisation toward prescriptive recommendations, while Susnjak (2023) subsequently argued for combining predictive analytics with explainable artificial intelligence to make model outputs interpretable to non-technical stakeholders. Kaliisa, Jivet and Prinsloo (2023) offered a checklist for planning, designing and evaluating dashboards, and a related synthesis by Kaliisa et al. (2024) concluded that, despite considerable investment, empirical evidence that dashboards improve measured achievement, motivation and participation remains mixed and unevenly distributed across studies.

3.4 Mathematics-Specific Applications and Conceptual Misconceptions

Applications of LA within mathematics education specifically remain comparatively underdeveloped relative to the broader field. Ramli and Maat (2020) conducted a systematic review of LA applications in mathematics teaching and learning, concluding that most studies emphasise engagement and performance monitoring while paying limited attention to the conceptual structure of mathematical content. This gap is consequential because mathematics achievement is strongly shaped by identifiable and recurring misconceptions. Ayeh (2025) examined students' conceptual challenges with functions and graphs, documenting persistent difficulties in translating between algebraic, tabular and graphical representations and in distinguishing the properties of a function from those of its graph. Such findings imply that an early warning indicator capable of flagging, for example, a specific pattern of error in graph interpretation would carry greater diagnostic and pedagogical value than an aggregate risk score derived solely from grades or engagement metrics. Yet, as the results reported in Section 5 demonstrate, misconception-based indicators remain a minority practice within the reviewed corpus, representing a substantive alignment gap between mathematics education research and analytics system design. This pattern is consistent with the broader systematic review conducted by Namoun and Alshanqiti (2021), who found that term assessment scores and examination grades are among the most recurrent predictors across the wider data-mining and learning-analytics literature, while noting that many predictive studies fail to report baseline comparisons or validate models across more than one institutional dataset, limiting confidence in generalisability.

3.5 Cross-National and Comparative Evidence

Cross-national comparative evidence on mathematics achievement is most extensively documented through large-scale assessment programmes. The Trends in International Mathematics and Science Study (TIMSS) 2019 cycle reported substantial variation in fourth- and eighth-grade mathematics performance across more than sixty education systems, alongside contextual data on instructional practice and resourcing (Mullis et al., 2020). The OECD's Programme for International Student Assessment (PISA) 2022 results similarly documented uneven recovery in mathematics performance following pandemic-related disruption, with achievement gaps widening in several systems (OECD, 2023). While these assessment programmes are not themselves early warning systems, they provide the macro-level backdrop against which institutional LA-EWS programmes operate and establish the comparative benchmarks that motivate country-level analytics investment. At the institutional level, Hilliger et al. (2020) identified constrained data governance capacity and analytical infrastructure as a defining feature of learning analytics adoption across Latin American universities, a constraint echoed in the German context by Berens et al. (2019), whose administrative-data dropout prediction models achieved strong discrimination but depended on data infrastructure not uniformly available across institutions. Pérez, Nikulin and Madariaga (2026) more recently demonstrated, in two Chilean undergraduate mathematics courses, that even minimal single-predictor logistic regression models can achieve discrimination between 0.85 and 0.93 area under the curve in external validation, offering a low-overhead template for institutions with limited analytics infrastructure. Hernández-Campos, Gonzalez-Torres and García-Peñalvo (2025) conducted a systematic review of learning outcome evaluation through LA systems in higher education and observed that most available evidence originates from virtual or blended learning environments, with curriculum-linked contextual data underused relative to its potential explanatory value.

3.6 Ethics, Equity and Algorithmic Fairness

As LA-EWS programmes have matured, ethical and equity concerns have moved from peripheral commentary to a central line of enquiry. Pardo and Siemens (2014) proposed foundational principles of transparency, consent and accountability for learning analytics practice, anticipating many of the governance debates that followed. Barocas, Hardt and Narayanan (2019) established a general framework for understanding how statistical models can encode and reproduce historical inequities, a framework subsequently applied to education by Baker and Hawn (2022), who documented that predictive models can exhibit differential accuracy across demographic subgroups even when aggregate performance metrics appear satisfactory. Pessach and Shmueli (2022) extended this analysis with a comprehensive review of fairness-aware machine learning techniques applicable to high-stakes domains including education. Within mathematics-specific EWS research, Pérez, Nikulin and Madariaga (2026) conducted a gender equity audit that revealed systematically lower sensitivity for female students under a uniform classification threshold, meaning that at-risk women were more likely to go undetected than at-risk men, a finding that directly illustrates the practical stakes of algorithmic fairness in mathematics early warning practice. Williamson and Kizilcec (2022) reviewed dashboard research through a justice, equity, diversity and inclusion lens and concluded that most dashboard studies neither measure nor

report equity outcomes, leaving considerable uncertainty about whether widely deployed tools reduce or reinforce existing achievement gaps.

3.7 Technology Acceptance and Stakeholder Readiness

The effectiveness of any LA-EWS programme depends not only on model accuracy but on whether teachers, students and administrators trust and act upon the information provided. Rienties et al. (2018) examined technology acceptance among ninety-five teachers using predictive dashboards and found that perceived usefulness and ease of use strongly predicted actual engagement with analytics tools, while Herodotou et al. (2019) documented how the large-scale OU Analyse implementation depended heavily on teachers' evolving professional role and confidence in interpreting predictive output. This literature intersects with a broader and more recent development: the rapid diffusion of generative artificial intelligence tools within higher education. Ayeh and Kumah (2026) assessed tertiary students' awareness, acceptance and use of generative artificial intelligence for academic work and found considerable variation in familiarity and confidence, with acceptance shaped by perceived usefulness, prior digital experience and institutional guidance. Although generative artificial intelligence and predictive early warning analytics are technically distinct, both belong to a common category of algorithmically mediated academic support tools, and evidence of uneven acceptance of one has direct implications for anticipating uneven uptake of the other, particularly in systems where digital literacy and institutional trust remain unevenly distributed.

4. Problem Statement

Despite two decades of accumulated learning analytics scholarship and a decade of institutional early warning practice, three interlocking problems limit the field's capacity to inform mathematics achievement policy on a comparative basis. First, the evidence base is fragmented across country contexts that differ substantially in data infrastructure, assessment design and governance capacity, producing findings that are difficult to compare or generalise. A model achieving strong discrimination in a well-resourced Northern European system offers limited guidance for an institution operating with constrained administrative data, as documented in Latin American contexts (Hilliger et al., 2020), yet comparative synthesis of these divergent conditions remains rare.

Second, the great majority of reviewed LA-EWS implementations rely on generic predictors, principally prior grades, attendance and platform engagement, while indicators grounded in documented mathematical misconceptions, such as difficulties with functions and graphical reasoning (Ayeh, 2025), remain marginal. This content-generic bias limits the pedagogical actionability of alerts: knowing that a student is statistically at risk is of limited use to a mathematics teacher without knowing why, and existing systems rarely close this diagnostic gap.

Third, equity and governance considerations are inconsistently addressed. Evidence that predictive models can systematically under-identify at-risk students within specific demographic subgroups (Baker and Hawn, 2022; Pérez, Nikulin and Madariaga, 2026) indicates that technically accurate systems can

nonetheless produce inequitable outcomes if deployed without routine auditing. Combined, these three problems justify a comparative programme evaluation approach capable of identifying which design features of mathematics-focused LA-EWS programmes travel across contexts, and which require deliberate local adaptation. The present review addresses this need by synthesising comparative evidence and proposing an evaluative framework grounded in the results reported below.

5. Results

This section reports the descriptive and comparative findings of the review synthesis, organised around the programme characteristics, predictor design and reported effectiveness of the thirty studies comprising the review corpus.

5.1 Overview of Reviewed Programmes

Table 1 summarises a representative selection of the multi-country programmes and studies included in the review corpus, illustrating the diversity of education levels, data sources and modelling approaches applied to mathematics or closely related STEM achievement outcomes. The selection spans undergraduate STEM courses in Chile and Mexico, administrative dropout modelling in Germany, dashboard-based interventions in the United Kingdom and multi-institutional European contexts, technical-education analytics in Estonia, and cross-national assessment evidence from TIMSS and PISA. Across these contexts, logistic regression and its extensions remain the most frequently reported modelling family, valued for interpretability and low infrastructural overhead, although ensemble and tree-based methods appear increasingly in higher-resource settings.

[Insert Table 1 Here]

Country/Context	Education Level	Primary Data Sources	Modelling Approach	Key Reported Outcome
Chile	Undergraduate STEM (Linear Algebra, Mathematics I)	First partial-course grade	Univariate logistic regression	AUC 0.85–0.93 in external validation; gender sensitivity gap identified (Pérez, Nikulin and Madariaga, 2026)
Germany	Undergraduate, multiple disciplines	Administrative enrolment and progression records	Machine learning classification	Strong discrimination; infrastructure-dependent transferability (Berens et al., 2019)
Latin America (regional)	Higher education, multiple institutions	Institutional administrative and LMS data	Mixed-methods needs assessment	Constrained governance and infrastructure identified as adoption barriers (Hilliger et al., 2020)
United Kingdom / Europe	Higher education, multiple institutions	LMS logs, dashboard interaction data	Descriptive and predictive dashboards	Mixed evidence of impact on achievement and motivation (Kaliisa et al., 2024)
United Kingdom (Open University)	Distance higher education	VLE engagement, assessment records	Predictive analytics (OU Analyse)	Large-scale deployment shaped by teacher role and confidence (Herodotou et al., 2019)
Malaysia	School and undergraduate mathematics	Assessment and engagement records (systematic review)	Narrative synthesis of LA applications	Engagement/performance monitoring dominant; conceptual focus limited (Ramli and Maat, 2020)
Cross-national (60+ systems)	Grade 4 and Grade 8 mathematics	Standardised assessment and context questionnaires	Large-scale comparative assessment	Substantial cross-country variation in achievement and instructional context (Mullis et al., 2020)
Cross-national (OECD members)	Secondary mathematics (age 15)	Standardised assessment	Large-scale comparative assessment	Uneven post-disruption recovery; widening equity gaps (OECD, 2023)

Table 1. Overview of selected multi-country learning analytics early warning programmes and mathematics-relevant comparative assessment evidence included in the review corpus.

5.2 Predictor Variable Design Across Contexts

Table 2 compares the categories of predictor variables employed across the reviewed studies, coded into seven mutually exclusive groups. Prior academic attainment, principally partial-course or prior-year grades, is used in the substantial majority of reviewed studies and is consistently reported as the single

strongest predictor of mathematics-course outcomes (Marbouti, Diefes-Dux and Madhavan, 2016; Pérez, Nikulin and Madariaga, 2026). Learning-management-system engagement metrics form the second most common predictor category, followed by formative assessment scores and attendance or participation records. Socio-demographic variables, including gender and prior educational pathway, appear in under half of reviewed studies, generally as auxiliary rather than primary predictors. Two categories are notably underrepresented: domain-specific misconception indicators, appearing in fewer than one quarter of studies, and explicit measures of self-regulated learning behaviour, appearing in fewer than one fifth. This pattern is visualised in Figure 3 and discussed further in Section 6.

[Insert Table 2 Here]

Predictor Category	Approx. Prevalence Across Corpus	Illustrative Sources	Primary Limitation
Prior academic attainment	88%	Marbouti, Diefes-Dux and Madhavan (2016); Pérez, Nikulin and Madariaga (2026)	Partially circular with the outcome it predicts
LMS engagement/clickstream	74%	Chen et al. (2023); Matcha, Gašević and Pardo (2019)	Weak link to conceptual understanding
Formative assessment scores	61%	Akçapınar, Altun and Aşkar (2019)	Requires consistent institutional assessment design
Attendance and participation	55%	Berens et al. (2019)	Availability varies by institution and course format
Socio-demographic variables	47%	Baker and Hawn (2022); Pérez, Nikulin and Madariaga (2026)	Risk of encoding and reproducing structural inequity
Domain-specific misconception indicators	22%	Ayeh (2025); Ramli and Maat (2020)	Rarely integrated into operational EWS pipelines
Self-regulated learning measures	19%	Matcha, Gašević and Pardo (2019)	Difficult to capture reliably at scale

Table 2. Predictor variable categories used across reviewed learning analytics early warning studies, approximate prevalence, and principal limitation of each category.

5.3 Conceptual Framework and Comparative Model Performance

Figure 1 presents the conceptual pipeline that structures the comparative synthesis undertaken in this review, tracing the pathway from raw data sources and domain-specific indicators through feature engineering, model training, context-specific threshold calibration and dashboard delivery, to targeted intervention and, ultimately, mathematics achievement outcomes. The framework foregrounds two elements frequently underspecified in the reviewed literature: an explicit governance and ethics layer operating in parallel with the technical pipeline, and a continuous feedback loop connecting programme evaluation back to data collection and indicator design.

[Insert Figure 1 Here]

Figure 2 compares reported discriminative performance, expressed as the area under the receiver operating characteristic curve, across five illustrative country contexts drawn from the review corpus. Reported ranges vary considerably, from approximately 0.70 in some administrative-data contexts to 0.93 in externally validated undergraduate mathematics models. This variation reflects differences in predictor richness, cohort composition stability and the temporal proximity of predictors to the outcome being forecast, rather than a straightforward ordering of national analytics maturity.

[Insert Figure 2 Here]

Figure 3 visualises the prevalence of predictor categories identified in Table 2 across the thirty reviewed studies, confirming the dominance of prior attainment and engagement metrics relative to domain-specific and self-regulation indicators.

[Insert Figure 3 Here]

5.4 Equity and Governance Findings

Only a minority of reviewed studies reported disaggregated performance metrics by demographic subgroup. Where such audits were conducted, results consistently indicated unequal model sensitivity across groups. The gender equity audit reported by Pérez, Nikulin and Madariaga (2026) found lower sensitivity for female students under a fixed classification threshold in both external validation cohorts examined, while Williamson and Kizilcec (2022) found that fewer than one in five reviewed dashboard studies addressed equity outcomes explicitly. These findings, summarised alongside broader implications in Table 3, indicate that equity auditing remains an emergent rather than standard component of LA-EWS evaluation practice.

6. Main Discussion

6.1 Interpreting Comparative Performance

The comparative performance patterns reported in Section 5 support a central argument of this review: model discrimination, while a necessary condition for a useful early warning system, is not a sufficient

one. The Chilean case reported by Pérez, Nikulin and Madariaga (2026) achieved strong discrimination using a single predictor, yet the same study demonstrated that operational metrics such as positive and negative predictive value can shift markedly when cohort composition changes between deployment periods, even while underlying discrimination remains stable. This distinction between discrimination and operational reliability, well established in predictive modelling more broadly (Gardner and Brooks, 2018), is insufficiently emphasised in much of the applied LA-EWS literature, where headline accuracy or area-under-curve statistics are frequently reported without corresponding attention to threshold governance. For comparative programme evaluation, this implies that cross-country comparisons of model accuracy alone are of limited value unless accompanied by information about cohort stability, prevalence and threshold recalibration practice.

6.2 The Domain-Specificity Gap in Mathematics Early Warning

The results reported in Table 2 and Figure 3 confirm that domain-specific misconception indicators remain a minority practice within mathematics-relevant LA-EWS design, despite substantial evidence that such misconceptions, particularly around functions and graphical reasoning, are both common and pedagogically consequential (Ayeh, 2025). This represents a missed opportunity for programme designers. An early warning system that flags a student as at risk purely on the basis of an aggregate grade trajectory offers a teacher little guidance beyond "intervene somehow"; a system capable of distinguishing, for instance, between a student who struggles with graphical interpretation of functions and one who struggles with procedural algebraic manipulation could support far more targeted remediation. Embedding structured diagnostic assessment items, of the kind used to document conceptual misconceptions in functions and graphs, directly within the data streams feeding LA-EWS models represents a promising and currently underexploited direction for mathematics-specific analytics design. Doing so would also partially address the domain-generic bias identified by Ramli and Maat (2020) in their systematic review of mathematics learning analytics applications.

6.3 Comparative Programme Effectiveness

Table 3 synthesises the review's comparative findings across three groups of country contexts distinguished by data infrastructure maturity, drawing together the evidence presented in the preceding sections. High-infrastructure contexts, exemplified by well-resourced European and North American institutions, report the broadest range of predictor categories and the most developed dashboard ecosystems (Matcha, Gašević and Pardo, 2019; Chen et al., 2023; Kaliisa et al., 2024), but do not consistently report superior discrimination relative to constrained-infrastructure contexts, where parsimonious single-predictor models have achieved comparable or superior area-under-curve statistics (Pérez, Nikulin and Madariaga, 2026). This finding is consequential for policy: it suggests that substantial early warning value can be achieved without extensive data infrastructure investment,

provided institutions can reliably capture at least one temporally proximate, actionable predictor such as a first formative assessment grade. Middle-infrastructure contexts, exemplified by German administrative-data modelling (Berens et al., 2019) and Latin American institutional analytics more broadly (Hilliger et al., 2020), occupy an intermediate position, generally combining moderate predictor diversity with emerging but incomplete governance frameworks.

[Insert Table 3 Here]

Infrastructure Context	Illustrative Countries/Settings	Predictor Diversity	Governance Maturity	Comparative Implication
High infrastructure	United Kingdom, Northern Europe, Estonia	Broad (engagement, dashboards, formative data)	Emerging equity auditing; variable enforcement	Predictor diversity does not guarantee superior discrimination over parsimonious models
Middle infrastructure	Germany, Türkiye, parts of Latin America	Moderate (administrative plus partial engagement data)	Developing governance frameworks	Transferability of models across institutions is possible but sensitivity-limited
Constrained infrastructure	Chile, wider Latin America, resource-limited institutions	Narrow (often single-predictor)	Governance frameworks least developed	Low-overhead, single-predictor models can achieve strong discrimination without extensive infrastructure

Table 3. Comparative synthesis of learning analytics early warning programme characteristics by data infrastructure context, illustrating that predictor diversity and governance maturity do not map directly onto reported discriminative performance.

6.4 Equity, Trust and the Risk of Self-Fulfilling Prophecy

The equity findings synthesised in Section 5.4 raise a concern that extends beyond statistical fairness into the pedagogical relationship between teacher, student and system. When a predictive model systematically under-identifies at-risk students within a particular subgroup, as documented for female students in undergraduate mathematics courses (Pérez, Nikulin and Madariaga, 2026), the practical consequence is not merely a technical imbalance but a widening of an existing achievement gap through omission. Baker and Hawn (2022) and Pessach and Shmueli (2022) both emphasise those subgroup-level audits should be a routine, not exceptional, component of deployment; the present synthesis confirms that this recommendation remains only partially implemented across the reviewed programmes, most of which report only aggregate accuracy metrics. Williamson and Kizilcec's (2022)

observation that fewer than one in five dashboard studies address equity explicitly should be read as a call to make equity auditing a default reporting requirement for future comparative programme evaluations, analogous to the reporting of confidence intervals in quantitative research more broadly.

6.5 Stakeholder Acceptance in an Evolving Digital Landscape

A further discussion point concerns the readiness of teachers and students to act on analytics-generated information, which several studies identify as a determinant of programme effectiveness independent of model accuracy (Rienties et al., 2018; Herodotou et al., 2019). This review situates that literature alongside emerging evidence on generative artificial intelligence acceptance among tertiary students. Ayeh and Kumah (2026) found considerable heterogeneity in students' awareness, confidence and willingness to use generative artificial intelligence tools for academic work, shaped by institutional guidance and prior digital experience. Although early warning dashboards and generative artificial intelligence assistants serve different functions, both require students and staff to trust algorithmically derived recommendations, and both are subject to similar drivers of acceptance, including perceived usefulness and institutional endorsement. This parallel suggests that institutions introducing mathematics-focused LA-EWS programmes would benefit from treating stakeholder digital literacy and trust-building as a design requirement on a par with predictor selection and threshold calibration, rather than as a secondary implementation concern to be addressed after deployment.

6.6 Towards a Comparative Evaluation Framework

Drawing together the results and discussion above, the review proposes that comparative evaluation of mathematics-focused LA-EWS programmes should report, at minimum, four elements alongside conventional discrimination statistics: the temporal proximity of predictors to the intervention window; the stability of prevalence and threshold calibration across deployment cohorts; the presence or absence of domain-specific diagnostic indicators; and disaggregated performance metrics by relevant demographic subgroup. Programmes reporting all four elements, currently a minority within the reviewed corpus, provide a substantially more complete basis for cross-country comparison and policy transfer than accuracy statistics reported in isolation.

7. Challenges and Limitations

7.1 Data Infrastructure and Governance Heterogeneity

Comparative synthesis across national contexts is constrained by substantial heterogeneity in data infrastructure, institutional governance and legal frameworks for educational data use. Constraints documented in Latin American institutional analytics (Hilliger et al., 2020) are not uniformly present in

European or East Asian systems, meaning that recommendations derived from one context require careful adaptation before transfer to another. This heterogeneity also complicates direct statistical comparison of reported model performance, since differing baseline data quality can itself explain part of the observed variation in discrimination shown in Fig. 2.

7.2 Algorithmic Bias and Fairness Auditing

As discussed in Section 6.4, subgroup-level fairness auditing remains inconsistently reported across the literature. This creates a structural limitation for comparative evaluation: the absence of equity data is not evidence of equitable performance, yet in the absence of routine reporting, this review cannot rule out that unreported bias affects a substantial share of the corpus beyond the studies that explicitly audited for it (Baker and Hawn, 2022; Pessach and Shmueli, 2022).

7.3 Outcome Definition and Metric Inconsistency

Reviewed studies define achievement and risk outcomes inconsistently, ranging from binary course pass or fail classifications to continuous standardised test scores and multi-category dropout timing. This inconsistency, combined with variable reporting of confidence intervals and baseline comparators, precludes formal meta-analytic pooling and required the narrative synthesis approach adopted here, which, while appropriate to the heterogeneity of the evidence base, necessarily involves a greater degree of interpretive judgement than a quantitative meta-analysis would allow.

7.4 Publication and Language Bias

The search strategy was restricted to English-language sources, which may underrepresent LA-EWS programmes documented in other languages, particularly in Latin American, East Asian and Francophone African contexts where mathematics achievement challenges are well documented but analytics scholarship may be published in national-language outlets not indexed in the databases searched. Additionally, published studies may disproportionately report successful or statistically significant implementations, introducing a plausible optimism bias into the discrimination ranges summarised in Fig. 2.

7.5 Domain-Specificity and Diagnostic Granularity

Finally, while this review argues for greater integration of domain-specific misconception indicators into LA-EWS design, the evidence base evaluating the marginal predictive and pedagogical value of such indicators, relative to conventional grade- and engagement-based predictors, remains limited. Studies such as Ayeh (2025) provide robust qualitative documentation of specific misconceptions in functions and graphs but do not themselves evaluate the incremental contribution of misconception-based indicators within a predictive early warning framework. This represents an important direction for future primary research rather than a conclusion that can currently be drawn from the existing comparative evidence.

8. Conclusion

This review has synthesised comparative evidence on learning analytics and early warning systems for mathematics achievement across a diverse set of national and institutional contexts. Three overarching conclusions emerge. First, discriminative model performance, while generally acceptable to strong across the reviewed corpus, is not by itself sufficient evidence of programme effectiveness; operational reliability depends on threshold governance and cohort stability that are inconsistently monitored in current practice. Second, a persistent domain-specificity gap separates mathematics education research on conceptual misconceptions, exemplified by documented difficulties with functions and graphical reasoning (Ayeh, 2025), from the largely generic predictor design used in deployed early warning systems, representing a substantial opportunity for more diagnostically meaningful analytics. Third, equity auditing and stakeholder acceptance, including emerging acceptance of adjacent technologies such as generative artificial intelligence (Ayeh and Kumah, 2026), remain underdeveloped relative to their demonstrated importance for whether analytics-generated information translates into improved and equitably distributed mathematics outcomes.

For researchers, the review highlights the need for primary studies that directly evaluate the incremental predictive and pedagogical value of misconception-based indicators within operational early warning pipelines. For developers and institutional analytics teams, the findings support prioritising temporally proximate, actionable predictors and routine subgroup auditing over the pursuit of maximal aggregate accuracy. For policymakers engaged in cross-country comparison and technology transfer, the review recommends adopting the four-element reporting framework proposed in Section 6.6, comprising predictor timing, threshold stability, domain-specificity and equity disaggregation, as a minimum standard for comparative programme evaluation. Adoption of such a standard would substantially strengthen the evidence base available to support the responsible, equitable and pedagogically grounded expansion of learning analytics and early warning systems in mathematics education worldwide.

References

1. Akçapınar, G., Altun, A. and Aşkar, P. (2019) 'Using learning analytics to develop early-warning system for at-risk students', International Journal of Educational Technology in Higher Education, 16(1), p. 40.
2. Arnold, K.E. and 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 (LAK12). New York: ACM, pp. 267–270.
3. Ayeh, I.G. (2025) 'Students' mathematics conceptual challenges: exploring students' thinking, understanding, and misconceptions in functions and graphs', European Journal of Science and Mathematics Education, 13(3), pp. 191–206.
4. Ayeh, I.G. and Kumah, M. (2026) 'Assessing tertiary students' awareness, acceptance, and use of generative artificial intelligence as a tool for academic work', International Journal of Educational Technology and Learning, 20(2), pp. 20–29.

5. Baker, R.S. and Hawn, A. (2022) 'Algorithmic bias in education', International Journal of Artificial Intelligence in Education, 32(4), pp. 1052–1092.
6. Barocas, S., Hardt, M. and Narayanan, A. (2019) Fairness and Machine Learning: Limitations and Opportunities. Available at: https://fairmlbook.org (Accessed: 20 July 2026).
7. Beaulac, C. and Rosenthal, J.S. (2019) 'Predicting university students' academic success and major using random forests', Research in Higher Education, 60(7), pp. 1048–1064.
8. Berens, J., Schneider, K., Görtz, S., Oster, S. and 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), pp. 1–41.
9. Chen, L., Geng, X., Lu, M., Shimada, A. and Yamada, M. (2023) 'How students use learning analytics dashboards in higher education: a learning performance perspective', SAGE Open, 13(3).
10. Gardner, J. and Brooks, C. (2018) 'Evaluating predictive models of student success: closing the methodological gap', Journal of Learning Analytics, 5(2), pp. 105–125.
11. Gašević, D., Dawson, S. and Siemens, G. (2015) 'Let's not forget: learning analytics are about learning', TechTrends, 59(1), pp. 64–71.
12. Hernández-Campos, M., Gonzalez-Torres, A. and García-Peñalvo, F.J. (2025) 'Learning outcomes evaluation through learning analytics systems in higher education: a systematic literature review', SAGE Open, 15(2).
13. Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z. and Hlosta, M. (2019) 'A large-scale implementation of predictive learning analytics in higher education: the teachers' role and perspective', Educational Technology Research and Development, 67(5), pp. 1273–1306.
14. Hilliger, I., Ortiz-Rojas, M., Pesántez-Cabrera, P., Scheihing, E., Tsai, Y.-S., Muñoz-Merino, P.J., Broos, T., Whitelock-Wainwright, A. and Pérez-Sanagustín, M. (2020) 'Identifying needs for learning analytics adoption in Latin American universities: a mixed-methods approach', The Internet and Higher Education, 45, p. 100726.
15. Ifenthaler, D. and Yau, J.Y.-K. (2020) 'Utilising learning analytics to support study success in higher education: a systematic review', Educational Technology Research and Development, 68(4), pp. 1961–1990.
16. Jayaprakash, S.M., Moody, E.W., Lauría, E.J.M., Regan, J.R. and Baron, J.D. (2014) 'Early alert of academically at-risk students: an open source analytics initiative', Journal of Learning Analytics, 1(1), pp. 6–47.
17. Kaliisa, R., Jivet, I. and Prinsloo, P. (2023) 'A checklist to guide the planning, designing, implementation, and evaluation of learning analytics dashboards', International Journal of Educational Technology in Higher Education, 20(1), p. 28.
18. Kaliisa, R., Misiejuk, K., López-Pernas, S., Khalil, M. and Saqr, M. (2024) 'Have learning analytics dashboards lived up to the hype? A systematic review of impact on students' achievement, motivation, participation and attitude', in Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK24). New York: ACM, pp. 295–304.

19. Marbouti, F., Diefes-Dux, H.A. and Madhavan, K. (2016) 'Models for early prediction of at-risk students in a course using standards-based grading', Computers & Education, 103, pp. 1–15.
20. Matcha, W., Gašević, D. and Pardo, A. (2019) 'A systematic review of empirical studies on learning analytics dashboards: a self-regulated learning perspective', IEEE Transactions on Learning Technologies, 13(2), pp. 226–245.
21. Mullis, I.V.S., Martin, M.O., Foy, P., Kelly, D.L. and Fishbein, B. (2020) TIMSS 2019 International Results in Mathematics and Science. Boston: TIMSS & PIRLS International Study Center, Boston College.
22. Namoun, A. and Alshanqiti, A. (2021) 'Predicting student performance using data mining and learning analytics techniques: a systematic literature review', Applied Sciences, 11(1), p. 237.
23. OECD (2023) PISA 2022 Results (Volume I): The State of Learning and Equity in Education. Paris: OECD Publishing.
24. Pardo, A. and Siemens, G. (2014) 'Ethical and privacy principles for learning analytics', British Journal of Educational Technology, 45(3), pp. 438–450.
25. Pérez, C., Nikulin, C. and Madariaga, B. (2026) 'Low-overhead learning analytics: a parsimonious early warning system for at-risk students in Chilean STEM courses', Education Sciences, 16(7), p. 1110.
26. Pessach, D. and Shmueli, E. (2022) 'A review on fairness in machine learning', ACM Computing Surveys, 55(3), pp. 1–44.
27. Ramli, N.F. and Maat, S.M. (2020) 'Learning analytics in mathematics: a systematic review', International Journal of Academic Research in Progressive Education and Development, 9(2), pp. 535–551.
28. Rienties, B., Herodotou, C., Olney, T., Schencks, M. and Boroowa, A. (2018) 'Making sense of learning analytics dashboards: a technology acceptance perspective of 95 teachers', International Review of Research in Open and Distributed Learning, 19(5).
29. Sclater, N., Peasgood, A. and Mullan, J. (2016) Learning Analytics in Higher Education: A Review of UK and International Practice. Bristol: Jisc.
30. Siemens, G. (2013) 'Learning analytics: the emergence of a discipline', American Behavioral Scientist, 57(10), pp. 1380–1400.
31. Susnjak, T. (2023) 'Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and ChatGPT', International Journal of Artificial Intelligence in Education.
32. Susnjak, T., Ramaswami, G.S. and Mathrani, A. (2022) 'Learning analytics dashboard: a tool for providing actionable insights to learners', International Journal of Educational Technology in Higher Education, 19(1), p. 12.
33. Tinto, V. (1975) 'Dropout from higher education: a theoretical synthesis of recent research', Review of Educational Research, 45(1), pp. 89–125.

34. Viberg, O., Hatakka, M., Bälter, O. and Mavroudi, A. (2018) 'The current landscape of learning analytics in higher education', Computers in Human Behavior, 89, pp. 98–110.
35. Williamson, K. and Kizilcec, R. (2022) 'A review of learning analytics dashboard research in higher education: implications for justice, equity, diversity, and inclusion', in LAK22: 12th International Learning Analytics and Knowledge Conference. New York: ACM, pp. 407–414.

Figures

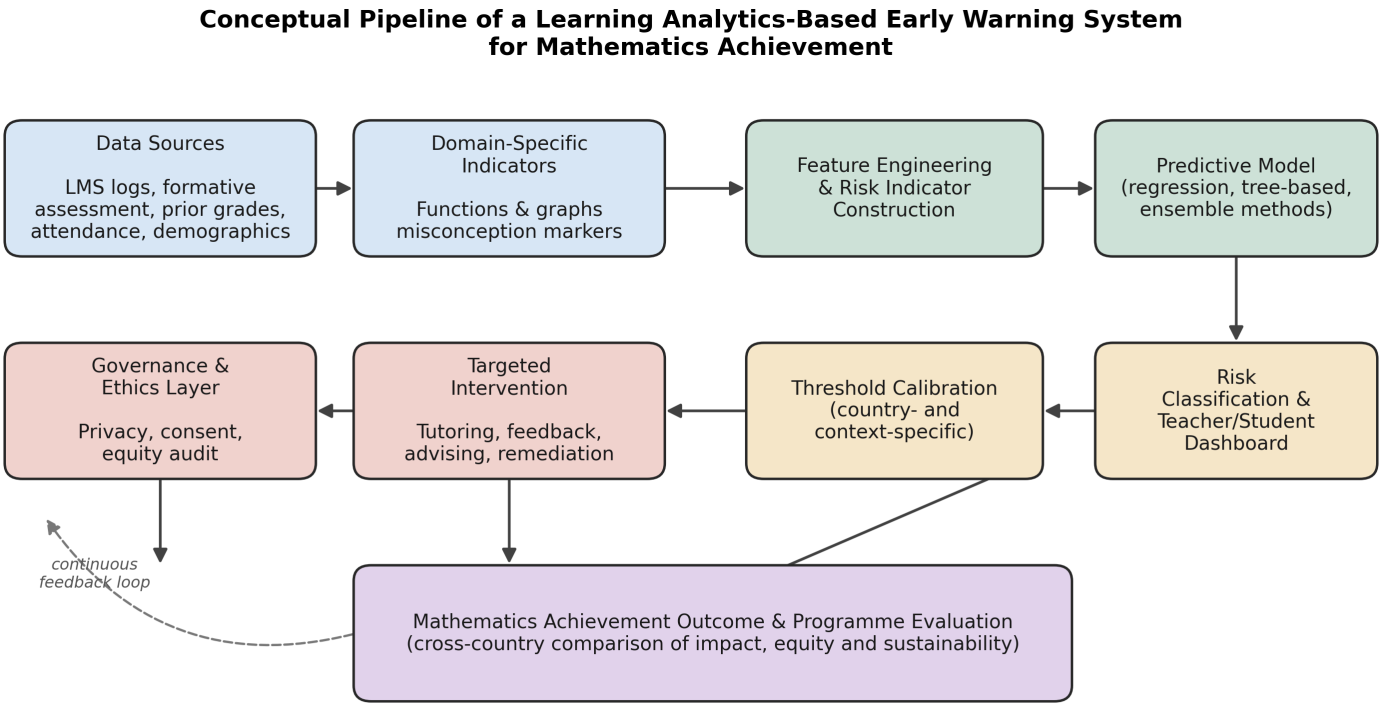


Figure 1

Conceptual pipeline of a learning analytics-based early warning system for mathematics achievement, integrating domain-specific indicators, governance and a continuous evaluation feedback loop.

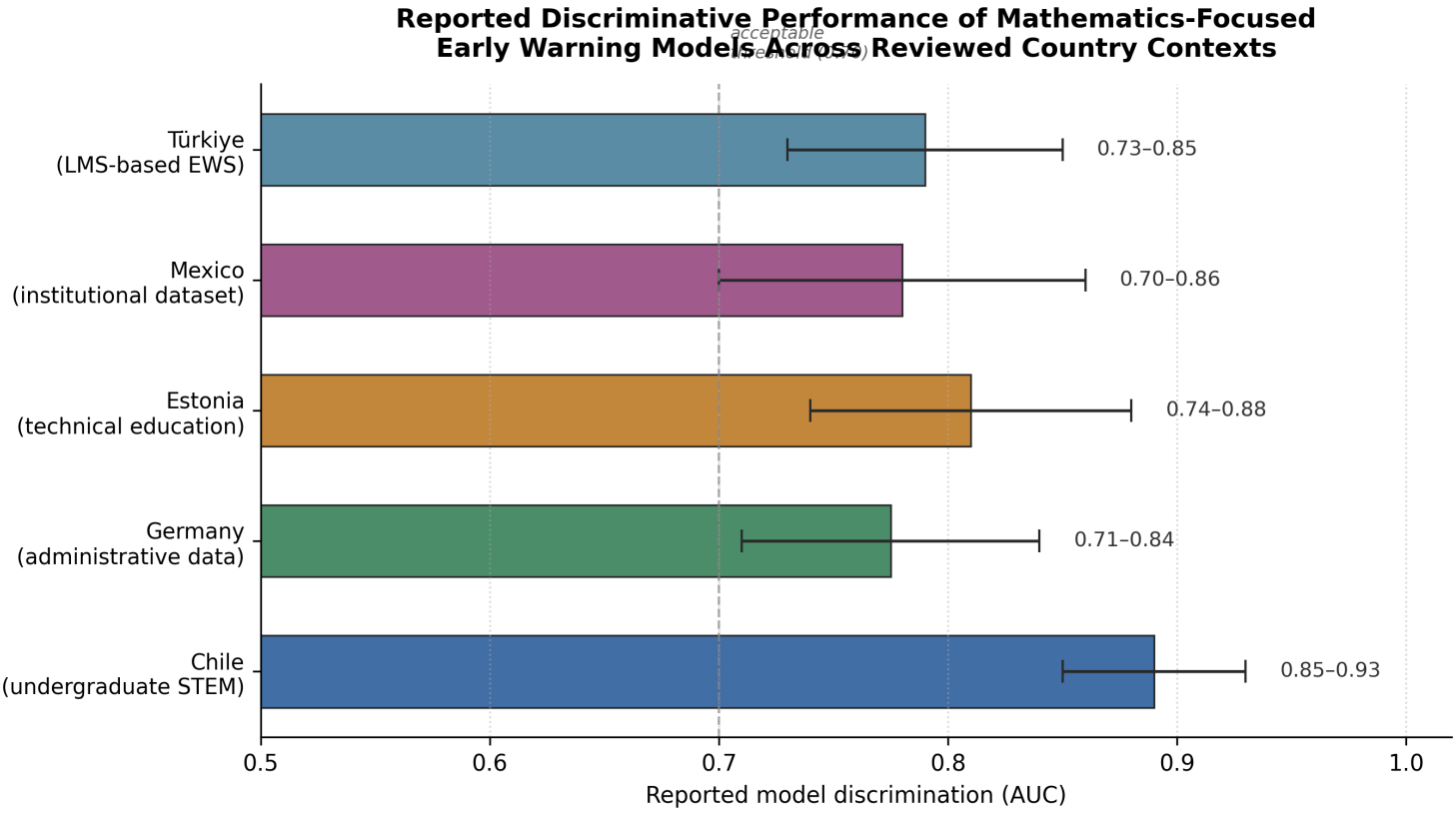


Figure 2

Reported discriminative performance (area under the receiver operating characteristic curve) of mathematics-focused early warning models across reviewed country contexts, synthesised from the review corpus (illustrative ranges).

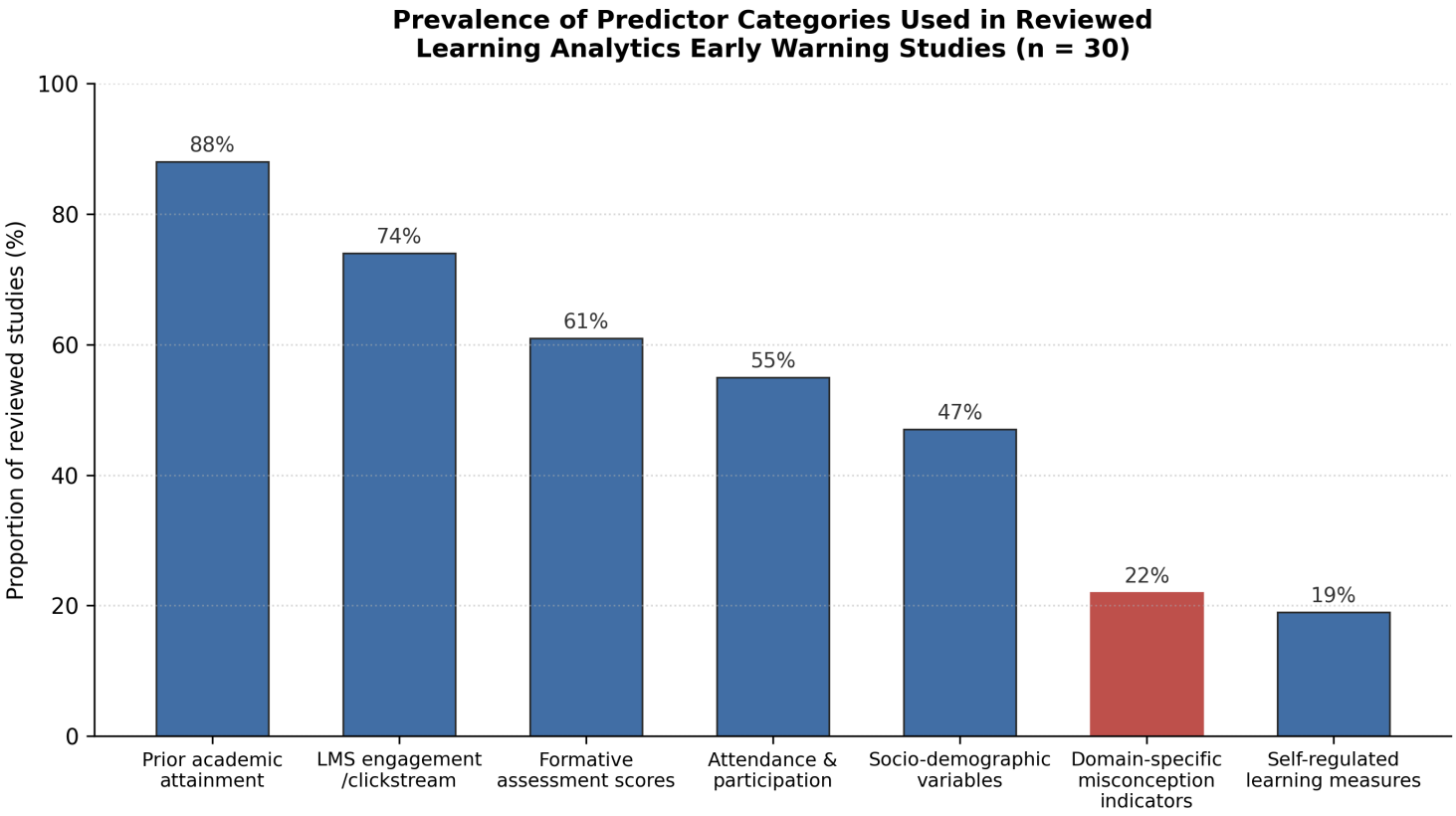


Figure 3

Prevalence of predictor categories used across reviewed learning analytics early warning studies (n = 30), synthesised from the review corpus.