Insights from a Five-Year Academic Analytics Observatory: Challenges and Achievements

Tiago Franco
CeDRI and SusTEC
Instituto Politécnico de Bragança
Bragança, Portugal
tiagofranco@ipb.pt

Paulo Alves
CeDRI and SusTEC
Instituto Politécnico de Bragança
Bragança, Portugal
palves@ipb.pt

José Rufino
CeDRI and SusTEC
Instituto Politécnico de Bragança
Bragança, Portugal
rufino@ipb.pt

Maria de Fátima Pacheco
CeDRI and SusTEC
Instituto Politécnico de Bragança
Bragança, Portugal
pacheco@ipb.pt

Nuno Ribeiro
UNIAG
Instituto Politécnico de Bragança
Bragança, Portugal
nunoa@ipb.pt

2025 International Symposium on Computers in Education (SIIE) | 979-8-3315-6257-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/SIIE68580.2025.11368366

***Abstract*—This paper presents a five-year case study of the development and operation of the Academic Success Observatory, an institutional platform designed to support data-driven academic management in higher education. The system combines automated data extraction modules, interactive dashboards, and a machine learning-based dropout prediction model. Drawing from three institutional databases, updated weekly, the platform enables continuous monitoring of academic indicators, report generation, and targeted support actions. The article discusses key technical and institutional challenges faced, as well as the benefits of integrating the platform into the academic decision-making process. The results highlight the system's potential to enhance decision-making, enable faster interventions, and foster a data-informed institutional culture.**

***Index Terms*—Academic Analytics, Learning Analytics, Dropout Prediction, Educational Dashboards, Higher Education**

I. Introduction

In recent decades, higher education institutions have increasingly accumulated large volumes of educational data, generated by academic management systems, virtual learning environments, and integrated administrative platforms. Despite the increasing digitalization of institutional processes, transforming data into effective actions to reduce student retention and promote academic success remains one of the key challenges in the educational sector [1].

To better understand this challenge, it is useful to consider the distinction between Academic Analytics (AA) and Learning Analytics (LA). According to George Siemens [2], Academic Analytics refers to the application of business intelligence techniques to support institutional-level decisions, such as enrollment management or resource allocation.

In contrast, Learning Analytics focuses on understanding and improving the learning process itself, by analyzing data about students, their behavior, and their interaction with educational environments. This distinction is essential for clarifying the different goals and stakeholders involved in the use of educational data, emphasizing the importance of initiatives that address not only administrative efficiency but also the quality and depth of the learning experience [3].

While the potential of analytics in education has been widely discussed, most initiatives documented in the literature remain confined to experimental settings, short-term projects, or operate without direct integration into institutional workflows. Even more rarely are there reports of platforms that remain operational over multiple years, evolving in response to ongoing technological and pedagogical changes [4].

In response to this scenario, the Polytechnic Institute of Bragança (IPB), a portuguese higher education institution, developed, in 2019, the Academic Success Observatory (ASO), an institutional platform aimed at the analytical management of students' academic trajectories. With a modular structure, it was designed to consolidate periodically updated data into a single environment, facilitating the work of academic and administrative managers, reducing dependence on multiple systems, and enabling greater agility in data analysis.

The ASO platform combines automated data extraction processes, interactive visualization through dashboards, access control mechanisms, and predictive modeling, allowing institutional managers to monitor academic indicators throughout the academic year without having to rely solely on traditional end-of-semester reports. A technical description of the platform's initial design and integration with IPB's internal systems was previously presented by Franco et al. [5], with a focus on the implementation of the system architecture.

The observatory was originally developed with a singular focus on dropout prediction, resembling other initiatives reported in the literature that failed to progress beyond the experimental or pilot stage. However, the incorporation of analytics functionalities aimed at day-to-day academic management proved decisive for expanding institutional adoption. This strategic reorientation allowed the platform to become part of the real workflow of institutional managers, consolidating its practical value and avoiding its perception as merely another isolated attempt to apply predictive models.

Over five years of continuous operation, the ASO platform has evolved from a technical solution to a decision support tool. Its incremental development, combined with continued use by education managers, has enabled the incorporation of technical improvements, regulatory compliance adjustments, and the institutional maturation of data-driven practices.

This article aims to report on this five-year experience, based on the development and operational trajectory of the platform. Through a descriptive and analytical approach, this paper discusses the main technical and institutional challenges encountered, the strategic decisions adopted, and the lessons learned. We hope to provide both practical insights and conceptual guidance for other institutions seeking to develop sustainable educational analytics initiatives.

The remainder of this article is organized as follows. Section II reviews related work on academic and learning analytics. Section III details the platform's architecture, data pipeline, and predictive model. Section IV presents operational performance and institutional outcomes. Section V concludes with final reflections and future directions.

II. Literature Review

Although the field of Academic Analytics (AA) and Learning Analytics (LA) has expanded rapidly in academic literature and institutional interest, evidence shows that its effective adoption in higher education institutions remains limited and fragmented. A systematic review conducted by Sønderlund et al. [6] examined 689 studies on analytics-based interventions and found that only 11 reported concrete interventions with measured impact, representing less than 2% of the reviewed cases. These findings highlight a disconnect between the theoretical enthusiasm surrounding LA and AA, and their practical effectiveness within institutional settings.

This conclusion is supported by Du et al. [7], whose meta-review showed that most published LA studies remain at the proof-of-concept stage, with limited integration into institutional decision-making processes. The authors emphasize that only 15% of the studies reviewed addressed strategic elements such as data governance policies, staff training, and alignment with academic management, all of which are essential for sustainable adoption.

The field of AA has also shown limitations in scale and impact. As highlighted by Llauró et al. [8], AA initiatives frequently face challenges related to fragmented data systems and departmental silos, which limit the effective integration of analytics into institutional decision-making and daily operational workflows. In a more recent study, Tsai et al. [9] similarly found that unclear data policies, siloed organizational structures, and weak communication between stakeholders are major factors limiting the adoption of analytics in large university settings.

In an attempt to identify conditions that favor effective institutional adoption, Márquez et al. [10] conducted a review of more than 100 studies and identified 14 critical success factors for institutional analytics initiatives. These include leadership engagement, system integration, faculty engagement, data governance structures, and collaboration between technical and pedagogical teams. The review showed that even among the few documented successful cases, no single framework addressed all these dimensions comprehensively and less than 10% implemented more than half of these factors, revealing the structural complexity involved in institutional-scale adoption.

In the domain of predictive analytics, Bacus and Cascaro [11] report that predictive LA has shown potential for improving academic success by enabling early identification of at-risk students and tailoring interventions more precisely. However, they also note that methodological inconsistencies and the lack of longitudinal evaluations limit the generalizability of current findings.

In terms of visualization and decision-making, dashboards have become central tools in both LA and AA. However, Paulsen and Lindsay [4] found that although there have been improvements in interface design and usability, most dashboards still offer primarily descriptive visualizations and lack integration with institutional decision-making workflows or pedagogical interventions. The authors argue that despite growing interest in learning-centered design, institutional usage remains underdeveloped.

Ethical and legal concerns have become increasingly prominent in both domains. Ifenthaler and Yau [12], along with Catasús et al. [13], conducted reviews showing that a significant proportion of studies fail to address essential elements such as informed consent, data transparency, and anonymization protocols, which are fundamental for compliance with regulatory frameworks like the General Data Protection Regulation (GDPR).

In summary, both Learning Analytics and Academic Analytics present valuable opportunities to improve educational quality and decision-making. However, as evidenced by several systematic reviews, their sustained and large-scale implementation remains the exception rather than the rule. Bridging this gap requires addressing not only technical and methodological challenges, but also institutional and ethical dimensions.

III. Academic Success Observatory

The Academic Success Observatory (ASO) is a digital platform developed to support data-informed academic management and early intervention in higher education. ASO integrates information from multiple academic subsystems into a unified, secure, and continuously updated environment. Its architecture was built to ensure autonomy for each institution while maintaining analytical consistency and scalability.

Rather than relying on direct queries to operational databases, the platform uses a scheduled and structured data pipeline, enabling reliable access to curated and transformed datasets. This approach not only reduces dependency on technical teams for ad hoc reporting, but also enables consistent monitoring of academic performance, student engagement, and dropout risk reduction over time.

A. System Architecture

The architectural design of the ASO observatory, illustrated in Figure 1, was initially introduced by Franco et al. [5] as part of a broader effort to promote sustainable data use in educational decision-making. The platform was designed to interface with existing systems already in use at higher education institutions, avoiding disruption while enabling advanced analytics through modular and secure infrastructure.

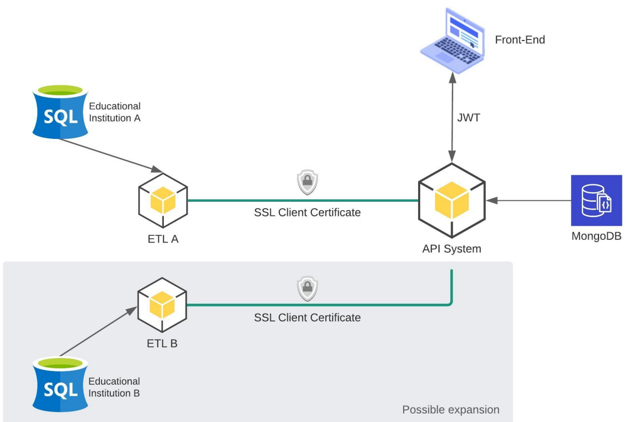

Fig. 1. ASO System Architecture [5].

At its core, the architecture relies on an Extract, Transform, Load (ETL) component responsible for collecting data from institutional academic systems. Each institution operates its own ETL module, installed locally and configured to extract relevant data directly from internal databases. The ETL component preprocesses the data by applying standardization, encryption of sensitive fields, and validation rules before transmitting it to a centralized API service.

The API system acts as the central integration point and is designed according to REST principles. It handles user management, task scheduling for data extraction, and execution of analytical and predictive modules. The processed data is stored in a MongoDB[1] database, which supports flexible, high-performance access for both statistical analysis and interactive dashboards.

A key architectural decision was to isolate the ETL module within each institution's infrastructure. This ensures that data extraction remains under institutional control and that only pseudonymized and preprocessed data is transmitted externally. This architecture supports compliance with privacy regulations such as the GDPR and national data protection laws, while also facilitating technical integration by minimizing interference with legacy systems.

B. Data Sources and Extraction Process

The ASO Observatory conducts its analyses based on periodic data extractions from three institutional databases: (i) *academic database*, which holds information on courses, subjects, students, and professors; (ii) *attendance database*, used to register classroom attendance; and (iii) *learning management system (LMS) database*, which stores records of student access and interaction with educational content.

The extractions are managed as automated tasks, each with a frequency determined by the nature of the data involved. Scheduling and execution are handled by the platform's API, which validates the data received, ensures extraction integrity, and triggers retries in case of temporary failure.

The extraction tasks are listed below:

1) **Academic Courses:** contains all active courses, including degree type, abbreviation, and duration.
2) **Academic Subjects:** lists curricular units offered in the academic year, including name, workload, and their associations with courses and departments.
3) **Students by Courses:** serves as the main student registry, recording each student's enrollment in their primary academic program, current semester, entry type, and history of completed and failed subjects.
4) **Students by Subjects:** identifies students enrolled in each subject, by course, and includes information such as final grades, assessment status, and whether the subject was previously attempted.
5) **Dropped-out Students:** lists students categorized as having dropped out, defined as those who were enrolled in the previous year but did not enroll in the following year and are not listed as graduates.
6) **Weekly Updates:** includes attendance records per class and LMS logs of access and interaction.

Tasks 1 through 5 involve relatively stable data and require the specification of the academic year to be extracted. These tasks are typically executed at the end of each semester. Task 6 is executed weekly and is the most complex in terms of volume and processing, as it integrates data from multiple sources and supports dynamic indicators used.

All tasks perform joins and aggregations across the three databases, generating datasets optimized for rapid access and visualization. The resulting structures are flattened and, in some cases, contain intentional redundancies to ensure high API performance in responding to front-end queries, regardless of the availability of institutional systems.

Extracted data does not include personal information such as names, emails, or contact details. Each student identifier is subjected to a reversible pseudonymization process at the time of extraction. Whenever it is necessary to access sensitive data, including basic requests such as listing enrolled students by course, the platform issues a controlled, on-demand request to the ETL module. These requests are logged by the API, including the user, timestamp, and type of data accessed. This audit mechanism is essential to ensure compliance with security policies and applicable data protection legislation, promoting transparency and responsible use of academic data.

C. Student Dropout Prediction

The student dropout prediction module within ASO was developed with the goal of identifying, in anticipation, students

[1] https://www.mongodb.com/

with a higher probability of dropping out. Its implementation followed a supervised machine learning approach, allowing historical patterns of academic behavior to be used in estimating, based on current data, the chance of dropout.

The model is based on historical academic data extracted through the tasks previously described. The variables used in training include performance indicators (such as the number of failed subjects, overall average grade, and percentage of credits completed), engagement metrics (weekly attendance and LMS activity), and institutional characteristics (such as degree program and type of enrollment). The algorithm adopted is a Random Forest classifier, implemented using the Scikit-Learn[2] library. It was selected for its robustness in handling heterogeneous variables, its capacity to capture complex interactions between features, and its higher interpretability when compared to other machine learning algorithms.

Each week, the model is retrained using historical records from the previous four academic years, considering only the data that would have been available up to the corresponding week in each year. This constraint ensures that the model does not use information from the future relative to the prediction point. After training, the model classifies all active students as either dropout-risky or not, based on their current data.

The dropout probability is defined as the proportion of weeks in which a student receives a dropout-risky classification. For example, if a student is classified as a dropout-risky every week, the dropout probability is 100%. If classified in only half of the weeks, the value is 50%. This cumulative indicator is then used to generate a priority ranking that supports follow-up actions carried out by specialized professionals. The platform includes a dedicated interface for student contacts, which displays this probability and supports the management of phone calls, allowing professionals to investigate issues reported by students and direct them to the appropriate services.

IV. Performance and Institutional Insights

This section presents operational evidence derived from five years of continuous use of the Academic Success Observatory at IPB. The focus is on the performance of the data pipeline and predictive processing, as well as how platform design choices improved the reliability, responsiveness, and scalability of analytics across institutional contexts.

It is important to note that the figures included in this section present illustrative subsets of institutional data and do not reflect the actual numbers of the institution, in order to preserve confidentiality.

A. Weekly Processing Time and Data Scope Optimization

The core of the platform's analytical process is a weekly execution cycle, aligned with the structure of the academic calendar. Although initial implementations of the system considered all 52 calendar weeks for data collection and classification, this strategy proved inefficient due to the presence of long academic breaks and administrative gaps. A subsequent refinement restricted the cycle to 40 weeks per year (20 per semester). However, closer analysis of prediction outputs revealed that the final weeks of each semester, typically weeks 17 to 20, offered little variation compared to week 16, often introducing noise without improving model sensitivity.

The current system, therefore, calculates weekly indicators only for the 16 teaching weeks of each semester. Weeks dedicated to exams, retake periods, or breaks are now formally excluded from the prediction pipeline. This adjustment not only reduced the volume of redundant data, but also improved the performance of the dropout prediction model by better aligning the weeks between different school years.

As of the latest operational cycle, the system manages data for approximately 10,000 active students. The average processing time for one academic week is between 10 and 15 minutes, with weeks near the end of each semester typically requiring more time due to data accumulation. A breakdown of the time distribution for a standard 15-minute run includes approximately 2 minutes for ETL extraction, 6 minutes for data registration, aggregation and validation via the API, 4 minutes for model training and classification using the machine learning module, and 3 minutes to calculate the dropout probability based on weekly classifications.

Other extraction tasks are significantly faster, typically requiring less than one minute, with the exception of Task 2, which establishes the student-to-subject mapping and takes on average 3 minutes to complete. Processing an entire academic year using the current configuration requires approximately 7 hours including all necessary tasks.

B. Indicators, Dashboards and Institutional Use

The platform currently provides more than 40 visualizations, including charts, tables, and interactive reports, supporting institutional management across various levels. These views are structured by scopes, such as institution, school, course, department, and student, and access is granted according to user profiles. Four main profiles are available: institution president, school director, course coordinator, and department.

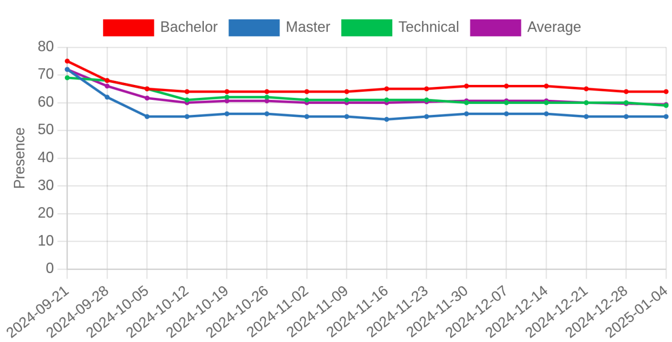

Fig. 2. Weekly evolution of average student attendance during the semester.

One of the most frequently accessed features is the home dashboard, which tracks the progression of four key indicators throughout the semester: average attendance (Figure 2),

[2]https://scikit-learn.org

number of enrollees, dropouts, and graduates. Among these, average attendance is particularly valuable for monitoring student engagement. The dashboard supports both aggregated and detailed analysis, making it a critical tool for early intervention and academic planning.

In addition to real-time metrics, the platform compiles historical indicators covering the last five academic years. These visualizations are frequently used in institutional reports and self-assessment procedures. Figure 3 shows the historical data of enrolled students, students who did not return in the following academic year (categorized as abandonment), and those who formally confirmed their withdrawal with the institution's administration (categorized as dropout). For the current academic year, it is not possible to identify students who will abandon their studies in the future, but only those whose dropout has already been confirmed.

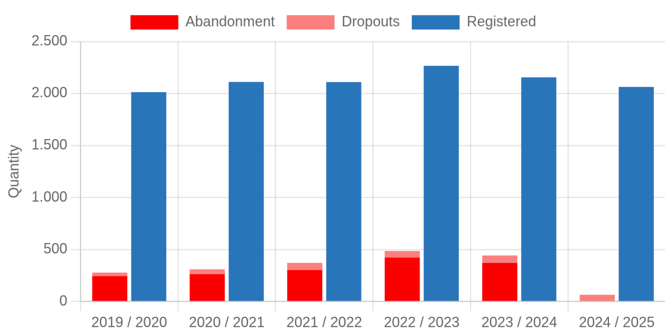

Fig. 3. Historical enrolled and dropped-out students over the last five years.

Another important set of indicators focuses on course-level performance. Figure 4 displays the percentage of students who were evaluated and passed, as well as the pass rate among those evaluated. These visualizations support the identification of high-risk subjects and help inform teaching strategies and resource allocation.

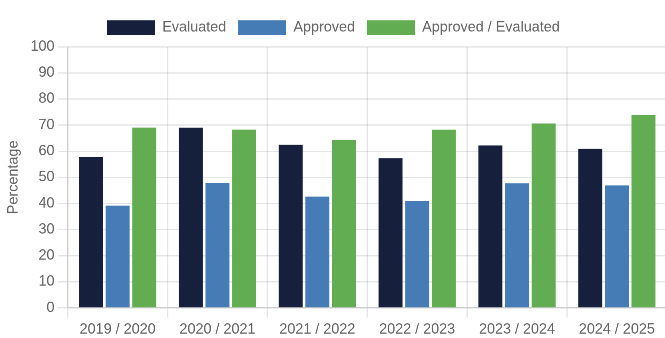

Fig. 4. Proportion of students evaluated and approved by subject.

By centralizing all these indicators in a unified platform with weekly automatic updates, the system has replaced previous workflows that depended on manually compiled reports from multiple sources. The result has been a significant increase in institutional agility, data reliability, and managerial engagement with analytics as part of routine academic monitoring.

C. Dropout Prediction and Follow-up Actions

The dropout prediction module, detailed earlier in the methodology, has been effectively integrated into the institution's student support routines. Rather than serving general academic management, the weekly predictions are made accessible exclusively to student support teams, who use them to prioritize outreach and guide intervention strategies.

Figure 5 shows a weekly comparison between the number of students predicted to drop out and those who actually left the institution by the end of the semester. This chart helps verify if the machine learning model behaves consistently with the previous year, indicating correct operation.

While not all predictions are confirmed, the results show an accuracy rate of nearly 90% for high-probability cases, with a recall of up to 60%. This suggests that almost all flagged students are experiencing significant challenges that hinder their academic.

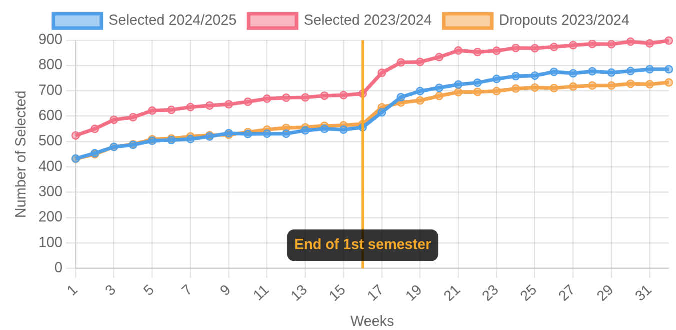

Fig. 5. Weekly predicted dropouts versus actual dropouts confirmed.

In addition to listing students at risk of dropout, the support team also has access to individual student reports, which include their course history, grades, attendance records, and a graphical timeline highlighting the weeks in which the student was classified as a potential dropout by the ML algorithm. In mid-2020, the ASO observatory was updated with a dedicated interface to support phone call tracking, marking the institution's first structured outreach campaign. Since then, follow-up calls have been conducted annually, reaching a total of 2,841 registered calls until now. Figure 6 shows the number of calls recorded per academic year.

As can be seen in Figure 6, the number of follow-up calls increased significantly in the most recent year. After observing promising results from earlier interventions focused on a single IPB school, the institution decided to expand the initiative. The most recent campaigns now include all IPB schools, allowing for broader outreach and reinforcing the institutional commitment to proactive student support.

V. Conclusions

Over five years of continuous operation, the Academic Success Observatory has evolved from a standalone technical solution into a strategic component of institutional academic management. Its ability to consolidate data from various systems, provide real-time analytics, and offer accessible visual-

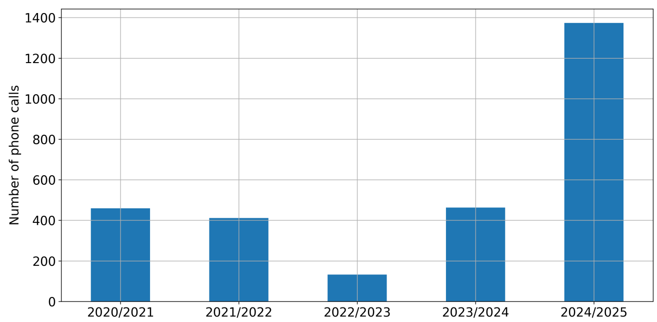

Fig. 6. Number of follow-up calls registered per semester since 2020.

izations has been central to increasing institutional engagement with data-informed practices.

Although implementing complex analytical mechanisms is far from trivial, the institutional benefits have been substantial. Beyond identifying students at risk, the platform has helped uncover structural issues affecting retention and academic success. Its integration into daily academic routines has contributed to the creation of new support programs and significantly increased referrals to specialized services.

Another important achievement has been the system's gradual maturation and, more recently, its full automation, allowing stable operation without the need for continuous technical intervention. This operational stability has strengthened user trust and opened the door for new use cases across different institutional sectors.

Future work will focus on expanding user profiles, adapting the platform for use in other institutions, and extending predictive analytics to other areas, such as academic performance and curricular progression. The experience described here demonstrates that sustainable and institutionally impactful analytics solutions are achievable, provided they are built upon continuous processes, cross-sector collaboration, and strategic vision.

VI. Acknowledgements

This work was supported by national funds: UID/05757 – Research Centre in Digitalization and Intelligent Robotics (CeDRI); SusTEC, LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020); and the project "Drop-in@IPB 2.0 – Caminhos para a Integração e Sucesso Académico dos Estudantes do Instituto Politécnico de Bragança", funded by RE-C06-i07 — Impulso Mais Digital, under the PRR – Recovery and Resilience Plan.

References

[1] M. Correa-Peralta, J. Vinueza-Martínez, and L. Castillo-Heredia, "Evolution, topics and relevant research methodologies in business intelligence and data analysis in the academic management of higher education institutions," *Results in Engineering*, vol. 25, p. 100 025, 2024. DOI: 10.1016/j.rineng.2024.103782.

[2] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," *EDUCAUSE review*, vol. 46, no. 5, p. 30, 2011.

[3] G. Siemens, "Learning analytics: The emergence of a discipline," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1380–1400, 2013. DOI: 10.1177/0002764213498851.

[4] L. Paulsen and E. Lindsay, "Learning analytics dashboards are increasingly becoming about learning and not just analytics - a systematic review," *Education and Information Technologies*, vol. 29, no. 11, pp. 14 279–14 308, 2024. DOI: 10.1007/s10639-023-12401-4.

[5] T. Franco, P. Alves, T. Pedrosa, M. J. Varanda Pereira, and J. Canão, "Implementation of big data analytics tool in a higher education institution," in *Trends and Applications in Information Systems and Technologies*, 2021, pp. 207–216. DOI: 10.1007/978-3-030-72657-7_20.

[6] A. Larrabee Sønderlund, E. Hughes, and J. Smith, "The efficacy of learning analytics interventions in higher education: A systematic review," *British Journal of Educational Technology*, vol. 50, no. 5, pp. 2594–2618, 2019. DOI: 10.1111/bjet.12720.

[7] X. Du, J. Yang, B. E. Shelton, J.-L. Hung, and M. Zhang, "A systematic meta-review and analysis of learning analytics research," *Behaviour & information technology*, vol. 40, no. 1, pp. 49–62, 2021. DOI: 10.1080/0144929X.2019.1669712.

[8] A. Llauró, D. Fonseca, E. Villegas, M. Aláez, and S. Romero, "Improvement of academic analytics processes through the identification of the main variables affecting early dropout of first-year students in technical degrees. a case study," *International Journal of Interactive Multimedia and Artificial Intelligence*, 2023. DOI: 10.9781/ijimai.2023.06.002.

[9] Y.-S. Tsai, S. Singh, M. Rakovic, L.-A. Lim, A. Roychoudhury, and D. Gasevic, "Charting design needs and strategic approaches for academic analytics systems through co-design," in *LAK22: 12th international learning analytics and knowledge conference*, 2022, pp. 381–391. DOI: 10.1145/3506860.3506939.

[10] L. Márquez, V. Henríquez, H. Chevreux, E. Scheihing, and J. Guerra, "Adoption of learning analytics in higher education institutions: A systematic literature review," *British Journal of Educational Technology*, vol. 55, no. 2, pp. 439–459, 2024. DOI: 10.1111/bjet.13385.

[11] J. A. Bacus and R. Cascaro, "Impact of predictive learning analytics in higher education: A systematic literature review," in *2024 13th International Conference on Educational and Information Technology (ICEIT)*, 2024, pp. 313–318. DOI: 10.1109/ICEIT61397.2024.10540840.

[12] D. Ifenthaler and J. Y.-K. Yau, "Utilising learning analytics to support study success in higher education: A systematic review," *Education Tech Research Dev*,

vol. 68, pp. 1961–1990, 2020. DOI: 10.1007/s11423-020-09788-z.

[13] M. G. Catasús, T. R. Fontanillas, J. E. Raffaghelli, and J. P. C. Martínez, "Collaborative learning analytics: Centring the ethical implications around teacher and student empowerment. a systematic review.," *Journal of Learning Analytics*, vol. 12, no. 1, pp. 201–214, 2025. DOI: 10.18608/jla.2025.8489.