

ORIGINAL ARTICLE

Learning Analytics in Higher Distance Education in Brazil: a bibliographical overview of national theses and dissertations (2009–2023)

*Fernanda Guinoza Matuda*¹
*Whaner Endo*²
*Patrícia Passos Gonçalves*³
*Dinah Feijó Capelo*⁴

ABSTRACT

This article provides a bibliographical overview of the adoption of Learning Analytics in Higher Distance Education in Brazil, based on a survey of national master's and doctoral research registered in the Brazilian Digital Library of Theses and Dissertations (BDTD) between 2009 and 2023. A total of 94 works were identified, of which 54 addressed Higher Education, and among these, 38 were directly related to distance and/or technology-mediated education contexts. The analysis highlights key concepts such as Educational Big Data and Educational Data Mining, as well as practical applications in areas like dropout prevention, teaching practices, and data analysis and visualization tools. The findings indicate that, despite the growing body of research, challenges persist regarding large-scale institutional adoption and the creation of dashboards and reports that support teachers and students in data-driven decision-making. It is concluded that integrating educational data with pedagogical approaches is essential to realizing the full potential of distance learning, underscoring the importance of privacy, accessibility, and data comprehension among all involved in the educational process.

Keywords: Learning Analytics. Distance Learning. Higher Education in Brazil. Educational Data. Educational Data Analysis.

1. Centro Universitário SENAC (fernanda.gmatuda@sp.senac.br)

2. Centro Universitário SENAC (whaner@gmail.com)

3. Centro Universitário SENAC (patricia.pagoncalves@sp.senac.br)

4. Centro Universitário SENAC (dinah.fcapelo@sp.senac.br)



Learning Analytics na Educação Superior a Distância no Brasil: um panorama bibliográfico de teses e dissertações nacionais (2009-2023)

RESUMO

Este artigo apresenta um panorama da utilização de Learning Analytics na Educação Superior a Distância no Brasil, com base em um levantamento de dissertações e teses registradas na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). Foram identificados 94 trabalhos, dos quais 54 abordavam o Ensino Superior e, entre estes, 38 relacionavam-se diretamente a contextos de educação a distância e/ou mediada por tecnologias. A análise destaca conceitos fundamentais, como Big Data Educacional e Educational Data Mining, bem como aplicações práticas em áreas como a prevenção de evasão, práticas de ensino, ferramentas de análise e visualização de dados. Os resultados indicam que, embora a pesquisa em Learning Analytics tenha crescido, ainda há desafios na adoção em larga escala e na disponibilização de dashboards e relatórios que apoiem diretamente docentes e estudantes na tomada de decisões. Conclui-se que a integração de dados educacionais e abordagens pedagógicas é fundamental para alavancar o potencial da EAD, ressaltando a importância da privacidade, acessibilidade e compreensão dos dados por todos os envolvidos no processo educacional.

Palavras-chave: Learning Analytics. Educação a distância. Ensino superior. Dados educacionais.

Learning Analytics en la Educación Superior a Distancia en Brasil: un panorama bibliográfico de tesis y disertaciones nacionales (2009–2023)

RESUMEN

Este artículo presenta un panorama bibliográfico sobre la adopción de Learning Analytics en la Educación Superior a Distancia en Brasil, basado en un relevamiento de trabajos de maestría y doctorado registrados en la Biblioteca Digital Brasileña de Tesis y Disertaciones (BDTD) entre 2009 y 2023. Se identificaron 94 investigaciones, de las cuales 54 abordaron la Educación Superior; de estas, 38 estaban directamente relacionadas con contextos de educación a distancia y/o mediada por tecnologías.



El análisis destaca conceptos fundamentales como Big Data Educativo y Minería de Datos Educativos, así como aplicaciones prácticas en áreas como la prevención del abandono, las prácticas de enseñanza y las herramientas de análisis y visualización de datos.

Los resultados indican que, a pesar del creciente número de estudios, persisten desafíos en la adopción institucional a gran escala y en la creación de paneles de control e informes que apoyen a docentes y estudiantes en la toma de decisiones basadas en datos. Se concluye que la integración de datos educativos con enfoques pedagógicos es esencial para aprovechar plenamente el potencial de la educación a distancia, enfatizando la importancia de la privacidad, la accesibilidad y la comprensión de los datos por parte de todos los involucrados en el proceso educativo.

Palabras clave: Learning Analytics. Educación a Distancia. Educación Superior en Brasil; Datos Educativos; Análisis de Datos Educativos.

1. Introduction

Data Science is a discipline grounded in research based on data collected for different purposes and analyzed in an exploratory manner, with the aim of generating discoveries and new knowledge. This type of research, according to Filatro (2021), follows a process that differs from the traditional approach (Hypothesis – Question – Data (primary) – Answer), as it involves the exploration of generated data with the goal of enhancing knowledge and supporting decision-making (Data (primary) – Exploration – Correlation – Insight).

As a subdiscipline of Data Science, Educational Data Science (EDS) emerged in the 2000s following a series of events and conferences focused on Educational Data Mining and, subsequently, on Learning Analytics (LA) beginning in 2011 (Filatro, 2021). EDS analyzes complex data and seeks to develop mathematical models to demonstrate findings and propose new approaches for educators and individuals involved in students' learning processes. Its research is therefore grounded in the collection, organization, and analysis of educational data and aims at the discovery and development of knowledge to support decision-making. It can thus be concluded that Learning Analytics is a field within Educational Data Science.

Lang et al. (2022) argue that the term is not easy to define; however, there are currently clearer understandings of what the field represents, largely due to the efforts of a scientific community that has come together around shared problems in order to advance the area. According to the authors, these efforts have taken place in recent years, particularly within higher education.

Educause, one of the most prominent nonprofit organizations in the field of education, had already indicated in the 2020 edition of its *Horizon Report* (Educause, 2020) that the use of Learning Analytics to improve student performance was a trend that should be monitored by educational institutions.

The current scenario of increasingly ubiquitous and pervasive education, marked by the use of virtual learning environments and mobile applications, offers a wide range of educational data sources, thereby strengthening projects related to the use of Learning Analytics with a focus on improving the quality of the teaching–learning process.

This article is one of the outcomes of the research project *Potentialities of Learning Analytics in Distance Education: Application of the CRISP-EDS Model in the Senac EAD Learning Environment*, which aims to understand the educational context and examine the process of data acquisition generated by the learning environment used by Senac EAD, while reflecting on how these data can be used to enhance student learning in this educational modality.

At this stage, the aim was to understand the landscape of academic research conducted on the topic in Brazil. To this end, a survey of master’s theses and doctoral dissertations registered in the Brazilian Digital Library of Theses and Dissertations (BDTD) was carried out, seeking to identify academic studies developed on the use of Learning Analytics in the context of distance higher education.

2. Methodology

A survey was conducted in the Brazilian Digital Library of Theses and Dissertations (BDTD), which compiles master’s theses and doctoral dissertations defended at Brazilian universities. The search was carried out on January 17, 2024, using the term **Learning Analytics** as the search descriptor.

In total, 160 records were identified and analyzed based on their content through a review of the titles and abstracts. This initial analysis revealed that 66 of the records did not address an education or teaching context and were therefore excluded from subsequent analyses. These studies focused on topics such as machine learning, the use of data for decision-making in public policy, among others.

Of the 94 studies related to the use of Learning Analytics in educational contexts, 31 are the results of doctoral research and 63 of master’s research, defended between 2009 and 2023. To identify the studies relevant to the scope of this bibliographic review, these works were classified according to educational level and theme/subject matter.

Table I – Studies Identified, Classified by Educational Level/Modality

Level or Educational Modality	Count
Basic Education	6
Corporate Education	2
Early Childhood Education	1
Elementary Education	5
Secondary Education	3
Higher Education	54
Technical Education	1
(blank)	22
Grand Total	94

Source: prepared by the authors

As can be observed from the analysis of Table I, 22 studies do not address a specific educational level. These are broader investigations that involve, for example, the use of Learning Analytics in the assessment of learning in general or of a specific subject or content area, but not necessarily within a formal educational context, such as the use of data generated in games to monitor learning (non-formal and/or informal education contexts). Among these studies, there are literature review investigations on the topic, such as the dissertation by Coelho (2019), which, among other objectives, sought to develop a systematic review of Deep Learning in the areas of Learning Analytics and Educational Data Mining.

The 54 studies related to higher education were categorized according to the teaching modality to which they referred. In this analysis, in order to identify distance education experiences—the focus of this bibliographic review—the following types of studies were considered: (I) those that make explicit reference to the distance education modality; (II) those that refer to virtual learning environments and/or teaching platforms; and (III) those that refer to remote education.

After this analysis, 38 of the 54 studies on higher education were categorized as distance education (EAD) and selected for the results analysis of this bibliographic review, together with the six studies that do not make specific reference to a teaching modality (face-to-face or distance learning) (Table II).

Table II – Studies on Higher Education by Teaching Modality

Modality	Count
Distance Education (EAD)	38
Face-to-Face	9
(blank)	6
Grand Total	54

Source: prepared by the authors

The selected studies were categorized based on their abstracts in order to map the academic productions according to subject matter and/or research objectives.

3. Results

As described in the previous section, 54 studies on the use of Learning Analytics in higher education were identified: nine studies related to face-to-face education, 39 associated with distance education contexts and/or technology-mediated education, and six that do not specify a particular teaching modality in their abstracts.

Of the six studies that do not address a specific teaching modality, four focus on the use of Learning Analytics resources for the analysis, prediction, and/or prevention of student dropout in higher education (Melo, 2016; Taborda, 2019; Vescovi, 2020; Guedes, 2021); one explores academic and socioeconomic data from an academic management system, analyzing how such data can support university actions (Rocha Neto, 2019); and one discusses challenges and controversies related to the use of Learning Analytics in university libraries, highlighting the need for further studies on the topic within the fields of Library and Information Science (Lago, 2020).

The 38 studies related to distance education contexts and/or technology-mediated education were analyzed through their titles and abstracts and organized into three categories:

Table III – Analyzed Studies by Category

Modality	Count
Educational Big Data	4
Teaching Practices	11
Learning Analysis	23
Grand Total	38

Source: prepared by the authors

The bibliographic analysis reveals an asymmetric distribution of production across thematic categories, making it possible to identify prevailing trends and significant gaps in the field. The category “Educational Big Data,” with only four studies, represents the smallest proportion of the works identified, although it points to promising paths for the use of large volumes of data for educational purposes. Next, the category “Teaching Practices,” with 11 studies, concentrates recent and still dispersed efforts to integrate Learning Analytics into everyday pedagogical practice, particularly with regard to the role of tutors, instructional mediation, and the promotion of interaction in virtual learning environments.

Finally, the category “Learning Analysis” emerges as the dominant thematic axis, with 23 studies (60% of the total), reflecting a significant emphasis in academic production on technical and quantitative approaches aimed at predicting student dropout, monitoring performance, and categorizing student profiles. This panorama highlights a convergence among authors regarding the potential of Learning Analytics as a tool for institutional management and decision-making support, while also revealing divergences concerning its pedagogical appropriation.

The interfaces between Learning Analytics and pedagogy, as well as the epistemological and political debates that could support a critical appropriation, remain marginally explored. This thematic asymmetry points to a significant formative gap, which limits the consolidation of Learning Analytics as a resource for transforming teaching–learning processes in distance education.

It is important to emphasize that each study has its own specific object, and some of them could be classified into more than one group. However, the categorization was carried out based on general themes, with the aim of grouping the studies by subject matter in order to achieve the objective of this exploratory research on academic studies related to Learning Analytics.

3.1 Educational Big Data and Learning Analytics

Based on the analysis of titles and abstracts, four studies were identified that address the structuring and use of Learning Analytics in higher education, exploring concepts such as Educational Big Data and Educational Data Mining.

Schneider (2022), in an exploratory study on maturity in the adoption of Learning Analytics in higher education, argues that the field of learning analytics has grown in recent years, but its adoption by institutions is still observed on a limited scale. Although most of the institutions surveyed by the author (located in the South and Southeast regions of Brazil) reported having some learning analytics initiative, there are still few institution-wide efforts. The author also notes that institutions have little or no difficulty in data acquisition and have already made progress in process-related areas, particularly with regard to privacy and ethics legislation. Schneider proposes an organizational maturity model that can assist in identifying the stage at which institutions are in terms of Learning Analytics adoption, serving as a basis for strategic actions. This study highlights the importance of an integrated institutional approach that takes into account not only technological infrastructure but also cultural and formative aspects, such as faculty development and data culture.

With regard to data collection, Biagiotti (2021) conceptualizes virtual learning environments as Educational Big Data, since they store large volumes of diverse data generated at high speed. Drawing on the experience of a MOOC (Massive Open Online Course), the author proposes a conceptual Learning Analytics framework to guide faculty and administrators in data management and use. According to Biagiotti (2021), the analysis of Educational Big Data (EBD) represents “a natural evolution of educational research methods” (Daniel, 2015, as cited in Biagiotti, 2021, p. 107) and can be understood as another mode of educational inquiry and research involving the analysis of large and complex datasets. This proposal stands out for emphasizing the articulation between data and teaching practice, positioning Learning Analytics as a tool for pedagogical mediation rather than merely for control.

On this same topic, Mendes (2017) conducts an applied study of a Big Data architecture in the context of the Moodle learning environment and defines Educational Data Mining (EDM) as a field whose objective is “to understand student behavior and the environment in which learning occurs, with the aim of improving learning, either by enhancing teaching quality or by optimizing students’ absorption of instruction” (Mendes, 2017, p. 24). The author notes that, although there are some overlaps, the field of Learning Analytics, or Learning Analytics Knowledge (LAK), differs from EDM in certain respects, one of which concerns the process of knowledge discovery: while the former prioritizes human intervention in the visualization and interpretation of data, the latter focuses on knowledge discovery through computational data processing. The focus of this study lies on technical feasibility and the robustness of the analytical infrastructure, demonstrating concern with system compatibility and solution scalability, albeit without advancing into pedagogical discussions.

Souza (2017) also addresses the need for information that enables decision-making by faculty and institutions, which increasingly use learning management systems (LMS) to offer “online education.” The researcher contextualizes the study by identifying problems arising from the growing use of e-learning resources and points to Business Intelligence as a pathway for proposing solutions that support the development of a software architecture capable of integrating tools and enabling the organization and interpretation of data for pedagogical decision-making by institutions. Inspired by market intelligence practices, the study aligns Learning Analytics with a performance- and efficiency-oriented logic, suggesting its adoption as a tool for academic management.

Based on the reading of the articles, it can be observed that although each study has distinct emphases and objectives, it is possible to identify common elements that point to shared directions within the field. Research in this category shows strong convergence in recognizing the potential of Learning Analytics to transform educational practices through the analysis of large volumes of data. All of the analyzed works acknowledge the importance of technological infrastructure, data organization, and intelligent information visualization as essential conditions for the consolidation of Learning Analytics in higher education. This convergence also extends to the appreciation of elements such as dashboards, frameworks, and integrated architectures to support decision-making processes at different institutional levels.

However, there are significant divergences regarding the focus of the contributions. Schneider (2022) and Biagiotti (2021) emphasize organizational and pedagogical dimensions, proposing maturity models and data-informed practices to support teaching. In contrast, Mendes (2017) and Souza (2017) adopt a more technological and managerial perspective, prioritizing the structuring of analytical systems and operational efficiency.

This difference in emphasis reveals a latent tension in the field: on the one hand, Learning Analytics as a tool to support pedagogical management; on the other, as an extension of management models grounded in performance and productivity.

As a trend, there is a strengthening of Learning Analytics as an interdisciplinary field in consolidation, marked by the intersection of education, data science, and information management. However, there remains a low incidence of formative or critical approaches capable of integrating data with the pedagogical principles that underpin teaching and learning processes. This suggests that the field remains strongly anchored in instrumental paradigms, lacking deeper epistemological engagement and stronger articulation with faculty development, especially in distance education contexts.

3.2 Teaching Practices

Eleven studies were grouped under the category “Teaching Practices,” encompassing experiences of applying Learning Analytics (LA) in higher education with a focus on instructional mediation, collaborative interactions, and course management in virtual learning environments. The studies are concentrated around three main axes: “analysis of participants’ actions,” “delivery of distance education courses,” and “interaction and collaboration.”

Table IV – Teaching Practices (Subcategories)

Category	Count
Analysis of Participants’ Actions	3
Delivery of Distance Education Courses	4
Interaction and Collaboration	4
Grand Total	11

Source: prepared by the authors

Among the studies categorized as “analysis of participants’ actions” are investigations that examine how the actions of participants—such as tutors and instructors—can contribute to decision-making and to the development of improvements in technology-mediated teaching–learning processes.

Acknowledging the complexities and specificities of the distance education modality, Souza (2016) sought to develop a behavioral assessment tool aimed at identifying whether the participation of tutors in undergraduate distance education courses could affect student performance. Using data extracted from a virtual learning environment to compare tutors’ actions with students’ actions in the analyzed cohorts, the study allowed the author to conclude that “distance tutors influence students’ effective participation in distance education classes” (p. 100).

The findings also make it possible to suggest adjustments so that tutors can carry out interventions that may improve the behavioral performance of the cohorts, especially with regard to sustained effective participation throughout the course.

Similarly, Fontes (2017) also sought to understand how tutors influence students' effective participation in distance education courses. The study was conducted based on historical data from courses offered in the distance education modality, involving hundreds of tutors and thousands of students. As outcomes, the author presents a model capable of correlating tutors' behavioral attributes with student participation, as well as the implementation of a Moodle plugin that enables the visualization of these data, supporting improvements in tutoring practices and course management.

Martins Filho (2016) investigated how Design Thinking can serve as a creative approach to foster innovation in teaching practices, focusing on the creation of knowledge assets for teaching and learning. The author highlights the need for a context-sensitive, institutionally grounded approach and emphasizes the active role of institutional leadership in promoting educational innovation.

The three studies analyzed support the recognition of the active role of tutors and instructors in the quality of teaching-learning processes in distance education courses, particularly with regard to pedagogical mediation and the impact of interactions on student performance. Souza (2016) and Fontes (2017) are grounded in the same premise: tutors' effective participation can directly influence student engagement and learning trajectories. Both studies use data generated by virtual learning environments (VLEs) to propose tutoring evaluation models, employing behavioral indicators as a basis for adjustments in educational practice. Their findings point to the feasibility of using Learning Analytics as a tool for monitoring, feedback, and the enhancement of pedagogical actions. Fontes (2017), in particular, makes a practical contribution by proposing MONITUM, a Moodle plugin that enables the visualization of correlations between tutors' and students' behaviors, demonstrating the potential of Learning Analytics to support large-scale educational management.

Martins Filho (2016), in turn, broadens the approach by proposing Design Thinking as a creative and institutional model for innovation in teaching practice, shifting the focus from the individual analysis of tutors to a more systemic and organizational perspective. By emphasizing the role of institutional leadership in fostering environments conducive to the creation of "knowledge assets," the author introduces a strategic dimension that goes beyond the use of Learning Analytics as a mere monitoring instrument and points toward its integration into processes of pedagogical and curricular redesign.

The main point of divergence among the studies lies in the scale and scope of their proposals: while Souza and Fontes work with individual metrics and focus on direct interactions between tutors and students, Martins Filho proposes a broader redesign guided by creative methodologies and institutional engagement. This distinction reveals two complementary rationalities in tension within the field: on the one hand, Learning Analytics as an instrument for microbehavioral analysis; on the other, Learning Analytics as an integral component of educational innovation policies.

As a prevailing trend, there is an increasing recognition of tutors and instructors as central agents in data-informed pedagogical decision-making processes. The studies indicate a shift from the purely technical use of Learning Analytics toward more integrated approaches that link data to formative practices and learning management. However, the challenge remains to consolidate models that combine monitoring with reflective processes, acknowledge the complexity of pedagogical mediation in digital environments, and expand the involvement of professionals in the production and critical interpretation of educational data.

The four studies classified under the subcategory “Delivery of Distance Education Courses” are investigations that use Learning Analytics to analyze the use of learning environments in various contexts, sharing a focus on data analysis to understand user behavior and enhance learning experiences.

Mota (2012), in his dissertation, investigates the use of a website to promote university extension activities aimed at small rural producers and family farmers, employing distance education. Similarly, Viel (2013) explores the application of ubiquitous computing technologies in distance education environments, focusing on the creation of interactive multimedia objects. He uses Learning Analytics to monitor users’ interactions with these objects and identifies areas for improvement in educational presentations.

In another teaching–learning context, Silva (2019) evaluated a MOOC (Massive Open Online Course) designed to train dentists in the health care of patients with chronic diseases, offered by the Universidade Aberta do SUS (UNASUS). The study applied Learning Analytics to correlate demographic factors, engagement, and the use of ICTs with participants’ course completion and certification rates. In a similar approach, Pereira (2019) explores best practices for the effective use of virtual learning environment tools in the mediation of higher education courses, particularly in the preparation of academic assignments. This study employs Learning Analytics to analyze whether the identified best practices, when applied in the courses, influence students’ behavior within the virtual learning environment.

The studies mentioned above contribute to understanding how distance education and Learning Analytics can impact different sectors and target audiences, providing insights for course improvement planning and for addressing the specific needs of students in virtual learning environments.

These studies position Learning Analytics as a relevant tool for pedagogical mediation and the enhancement of learning experiences in digital contexts. However, this convergence does not imply homogeneity. Divergences are evident both in the empirical objects, which range from formal academic training to university extension processes, and in the target audiences and methodological approaches. Viel (2013) and Silva (2019) operate with quantitative models and usage metrics, characterized by a high degree of technicality, relying on statistical analyses and correlations (rather than necessarily establishing causality) to support their conclusions. In contrast, Mota (2012) and Pereira (2019) emphasize the social and formative dimension of instructional mediation, suggesting that the pedagogical interpretation of data should be sensitive to the particularities of individuals and contexts.

This diversity of approaches reveals that Learning Analytics, far from being a closed field, is configured as a contested arena where different technical, pedagogical, and political rationalities intersect in shaping the meanings and uses of data.

As a trend, a gradual shift can be observed from the use of Learning Analytics as a mechanism for monitoring and control toward more proactive approaches that seek to understand students' learning trajectories in a situated, contextualized, and intentionally pedagogical manner. Although still incipient, this shift signals an openness to integrating data with critical elements of the educational process, such as instructional design, teacher mediation, and active student engagement. Nevertheless, the field still lacks a more robust theoretical foundation that would allow for a critical appropriation of data, incorporating principles of emancipatory pedagogy, educational justice, and ethics in the production and interpretation of evidence. The recognition of learners' agency and the contextual interpretation of indicators remain central challenges for Learning Analytics to consolidate not merely as a technology but as a transformative educational practice.

Finally, the studies grouped under the subcategory "Interaction and Collaboration" make a significant contribution to understanding how Learning Analytics (LA) can be applied to the analysis of interactive behaviors in virtual learning environments, particularly in distance education contexts. What unites them is the recognition of interactions—especially those occurring in discussion forums, peer exchanges, and collaborative activities—as relevant indicators for analyzing engagement and the quality of learning experiences. All the authors, despite differing approaches, start from the premise that data extracted from VLEs contain strategic information for understanding behavioral patterns and guiding pedagogical practices.

Brandão (2018) states that interactions are a "fertile field for researchers in educational data mining and Learning Analytics" (p. 8) and uses data extracted from an undergraduate course offered in the Moodle environment to identify patterns of student engagement and procrastination behaviors.

In line with this research approach, Ferreira (2018) analyzes interactions in discussion forums, problematizing this resource, which is often recognized for its collaborative learning potential but is not always used effectively, as students tend to respond primarily to tutors. The author presents a data mining technique to extract Learning Analytics in order to analyze collaboration within forum posts, based on five "collaborative characteristics," and to measure the impacts on pedagogical mediation and student-to-student collaboration.

Mônego (2019) conducted a case study to evaluate student and instructor participation and interaction in an online collaborative setting within virtual learning environments, focusing on the initial training of teachers in an undergraduate program in Exact Sciences. The author used questionnaires administered to participants and activity logs recorded in the learning environment as data collection instruments.

Also within this group of studies, Holanda (2020) presents a Collaboration Framework for MOOCs designed to promote greater student interaction. Using Learning Analytics and peer recommendations within the virtual environment, the model provides a monitoring dashboard that allows instructors to track both individual and collective student performance.

The convergences among these studies are evident: all recognize the potential of Learning Analytics to reveal hidden aspects of interaction dynamics in virtual learning environments (VLEs) and, in some way, propose methods to make these data intelligible and useful for enhancing pedagogical mediation. The emphasis on collaboration, even when operationalized through different techniques, emerges as a shared formative value, and discussion forums are identified as central (albeit sometimes problematic) spaces for realizing this goal.

However, the divergences are evident in how collaboration is understood and operationalized. Brandão and Ferreira adopt more technical and quantitative approaches, focused on pattern modeling and algorithm development. In contrast, Mônico situates the analysis within a broader pedagogical context, incorporating the perspectives of the participants involved. Holanda, meanwhile, advances in the personalization of the collaborative experience by proposing an architecture that connects peer recommendations with collective performance analysis. The methodologies employed range from data mining, time series analysis, and network analysis to case studies, highlighting the epistemic diversity within the field.

As a prevailing trend, there is a growing effort to operationalize collaboration in virtual learning environments more effectively and intelligibly, both for instructors and for students themselves. The studies point to a movement toward greater sophistication in the analytical resources applied to educational interaction, highlighting the use of visual dashboards, recommendation algorithms, and customized metrics. However, the challenge remains to overcome the asymmetry between instructor mediation and student participation in discussion forums, requiring approaches that more fully integrate data with pedagogical intentionality. The pedagogical critique of the passivity of traditional forums and the centralization of the tutor's role is a latent issue that needs further exploration for Learning Analytics to also serve a transformative function in the pedagogical culture of distance education.

3.3. Learning Analysis

The studies categorized under “Learning Analysis” examine the application of Learning Analytics in distance education (DE) environments, though they differ in specific objectives and methodological strategies. Despite this diversity of approaches, important convergences can be identified around the use of Learning Analytics as a tool to support pedagogical decision-making, personalize instruction, and address longstanding challenges in the modality, such as student dropout and lack of engagement.

A first grouping includes studies focused on identifying and mitigating student dropout and failure. Portal (2016), Cambuzzi (2014), Andrade (2023), and Brito (2019) primarily explore dropout, emphasizing the importance of understanding and addressing the causes of attrition in distance education. The significance of this topic is underscored by the high dropout rates in DE, making it a recurring focus in the field. In a complementary approach, Ferreira (2016) proposes a predictive model aimed at identifying at-risk groups for course failure.

Although it also addresses issues of course continuity, Ferreira's study offers a distinct contribution by focusing on failure prediction based on analyses of academic behavior, thus broadening the scope beyond studies that concentrate exclusively on dropout.

A second strand highlights studies focused on constructing student profiles and conducting contextualized analyses of the learning process, such as those by Nunes (2019) and Brasil (2019). Nunes (2019) develops a student profile data model aimed at the use of Learning Analytics in online learning systems, enabling analysis geared toward decision-making in the educational context. Similarly addressing student profiling, Brasil (2019) employs Learning Analytics in an analytical process oriented toward instructors, providing a contextualized interpretation of student data based on socioeconomic and educational factors.

Taken together, these studies reflect the development of Learning Analytics as a tool to support distance education, offering alternatives for both preventive interventions and personalized adaptations, as well as contextualized data analyses, thereby expanding understanding of student behavior and performance in DE. These works differ from others by seeking a more nuanced and situated understanding of student behavior, incorporating socioeconomic data, learning trajectories, and learning styles into the use of Learning Analytics. This approach represents an important advance in addressing the limitations of exclusively quantitative analyses, moving toward a more integrative and critical logic in educational assessment.

Complementing this analysis, the third subcategory focuses on studies that apply Learning Analytics to the measurement, planning, and evaluation of pedagogical practices, particularly regarding teaching strategies and active methodologies. Luz (2017), Dias Junior (2019), Perini (2019), Martins (2023), and Ribeiro (2018) exemplify this trend by employing Learning Analytics to support diagnostic assessment, measure learning outcomes, and monitor the impact of specific instructional resources and courses, such as Brazilian Sign Language (Libras) or Physics instruction. The convergence within this group lies in valuing Learning Analytics as a tool for pedagogical planning, while divergences relate to the depth of didactic-curricular analysis and the ways in which data are operationalized by instructors.

A fourth group of authors focuses on studies that treat Learning Analytics as a tool for monitoring and (self-)regulation of learning, highlighting the active role of students in tracking their own progress, particularly through dashboards and visualization models.

In this regard, Santos (2020a) highlights that most Learning Analytics Dashboards (LADs) are designed for instructors, leaving a gap in the development of dashboards specifically for direct student use. The author tests a solution aimed at enabling students to monitor their own learning performance and behavior, fostering awareness and reflection on their progress. In a complementary approach, Santos (2019) explores adaptive learning, integrating it with pedagogical models that value students' prior knowledge. The study demonstrates that courses structured around mediated learning experiences can support learning regulation by adapting content according to prior knowledge and promoting a personalized experience.

This analysis aligns with Santos (2020b) in seeking strategies to enhance students' awareness of their own learning processes, while also introducing the dimension of adaptive instruction to provide more structured support in courses with or without tutor guidance.

Lacerda (2018), in turn, proposes a methodology that uses sequential mining algorithms to classify VLE usage patterns according to categories of self-regulated learning. By classifying these patterns, the author contributes to a more nuanced understanding of self-regulation in distance education, suggesting that sequential data analysis can reveal unexpected behaviors and provide insights for more targeted interventions.

Other studies focus on the creation of visualization dashboards and the use of classification algorithms, highlighting the growing role of Learning Analytics in optimizing teaching and learning in distance education. Melo (2023), for example, using data visualization techniques and explainable artificial intelligence, suggests that these tools can help identify student profiles, providing valuable insights into the learning process and enabling more targeted pedagogical interventions.

Zapparoli (2016) highlights the difficulty instructors face in obtaining a comprehensive view of student participation, emphasizing the need for a reporting system that allows a cross-sectional analysis of interactions in VLEs. His study is based on the integration of Learning Analytics into the Moodle VLE, proposing a tool that provides instructors with a more holistic view of student activities, with reports aggregated by courses and contexts. Similarly, Oliveira (2016) notes that VLEs such as Moodle offer reports on student performance but do not always address the real needs of instructors, stressing the importance of a model that meets teachers' demands for diagnosing learning difficulties.

In line with this type of study, Damasceno (2020), by exploring instructors' preferences, developed a model to support the analysis of educational data in VLEs, focusing on teachers' needs regarding data visualization. His research also highlights the lack of models that assist instructors in analyzing student activity records.

Among the main points of convergence, there is widespread recognition of Learning Analytics as a valuable strategy for anticipating dropout behaviors, supporting pedagogical planning, and developing more refined monitoring solutions. The emphasis on building dashboards, visualization models, and predictive algorithms reinforces the field's movement toward producing tools that are directly applicable to learning management systems.

The divergences, in turn, manifest both in the objects of study and in the methodological and epistemological approaches adopted. Some of the works still operate under a technicist and instrumental logic, focusing on performance and retention variables. In contrast, other authors have proposed approaches that are more sensitive to context, student diversity, and learner agency, incorporating qualitative data, adaptive models, and concerns related to critical pedagogical development.

As a prevailing trend, there is an increasing integration of Learning Analytics with principles of personalized instruction, teacher development, and contextualized pedagogical mediation. Learning analysis is no longer solely a matter of control; it increasingly informs strategies of care, support, and more human-centered, formative decision-making. Nevertheless, the challenge remains to consolidate an institutional culture that critically appropriates data and promotes its ethical, reflective, and pedagogically situated use within educational processes.

4. Final Considerations

The bibliographic survey identified studies that emphasize the importance of consolidating fundamental terms in the field of Learning Analytics, such as Big Data, Educational Big Data, Educational Data Mining (EDM), Learning Analytics (LA), and Learning Analytics Knowledge (LAK). These often overlapping concepts reflect the interdisciplinary nature of LA, which integrates knowledge from computing, data science, and education to transform data into actionable insights for learning.

The analysis conducted in this systematic review revealed that, although the field of Learning Analytics applied to Distance Education (DE) has been expanding in Brazil over the past two decades, its scientific output still shows an uneven distribution across subtopics, with a predominance of technical-quantitative approaches. The highest concentration of studies falls within the “Learning Analysis” category—particularly those focused on predicting dropout, academic performance, and student behavior—confirming the centrality of institutional metrics as the main driver of academic interest. In contrast, categories such as “Educational Big Data” and “Teaching Practices” remain marginal, highlighting the ongoing challenge of critically integrating data and pedagogy in educational contexts.

Based on the triangulation of objectives, methodologies, and contributions of the studies analyzed, significant convergences were identified, such as the recognition of Learning Analytics as a tool to enhance educational planning, personalize learning, and support institutional management. On the other hand, important divergences also emerge regarding the epistemological depth of the approaches, the formative intent, and the role attributed to teacher mediation in the use of data.

As a prevailing trend, there is a gradual shift of Learning Analytics from being merely a control tool toward its appropriation in models that are more sensitive to the educational context, aiming to integrate data with instructional design, pedagogical mediation, and student agency. Studies proposing dashboards for student use, self-regulation methodologies, interaction mining, and connections with active learning strategies signal the expansion of Learning Analytics as an emerging and consolidating interdisciplinary field.

A particularly significant finding concerns the low number of publications identified: only 94 studies over 14 years. Although quantitative, this figure carries substantial analytical weight, as it highlights the fragility of the field in Brazil and calls for critical reflection on the factors that have hindered its consolidation as a foundational component of teacher training, pedagogical innovation, and the formulation of public policies in distance education.

This numerical limitation is not explained solely by the relative novelty of the topic but points to persistent structural, formative, and epistemological barriers. Among these are the lack of technological infrastructure for large-scale analysis, limited teacher training focused on the critical use of educational data, and the absence of institutional policies that encourage experimentation and applied research with Learning Analytics. Added to this is the predominance of a technician and instrumental conception of LA, which often overlooks its ethical, formative, and contextual implications, making it difficult to effectively incorporate data into the reflective processes of pedagogical practice.

In this context, the small number of studies serves as a symptomatic indicator of multiple gaps—technical, political, pedagogical, and epistemological—that must be urgently addressed. By highlighting this scarcity, the present study not only maps what has already been produced but also delineates a research horizon that values the critical integration of Learning Analytics into educational practices, recognizing its transformative potential when guided by pedagogical principles, social commitment, and the enhancement of learner autonomy.

A fundamentally recurring aspect identified in the analysis of several studies is the need to make educational data more accessible to both teachers and students, enhancing its use for informed decision-making and personalized instruction. However, although the field has progressed in developing frameworks and analytical models, there remains a lack of studies exploring the practical applicability of these models in diverse educational settings and testing the effectiveness of Learning Analytics across different distance education contexts. Furthermore, greater investigation is needed into the creation of tools that allow teachers and students to analyze data intuitively, thereby promoting learner autonomy in the regulation of their own learning.

Thus, this article offers a critical contribution by outlining the state of the art of national academic production on Learning Analytics in distance education and highlights the urgent need to deepen the dialogue between data science and education, incorporating pedagogical, ethical, and epistemological foundations into the analysis and use of educational data. Future research could explore in greater depth the intersections between Learning Analytics and teacher training, Learning Analytics and curriculum design, as well as investigate how data-driven practices can be co-constructed with learners in collaborative, critical, and socially situated contexts.

REFERÊNCIAS

ANDRADE, Tiago Luís de. **Metodologias ativas integradas a um sistema de recomendação com suporte à mineração de dados educacionais e Learning Analytics para a mitigação de evasão da educação a distância**. 2023. Tese (Doutorado em Computação Aplicada) – Universidade do Vale do Rio dos Sinos, São Leopoldo, 2023.

BIAGIOTTI, Breno de Almeida. **Framework conceitual para análise de Big Data Educacional em MOOCs**. Tese (Doutorado em Engenharia e Gestão do Conhecimento). Universidade Federal de Santa Catarina, Florianópolis, 2021.

BRANDÃO, José Orlando da Silva. **Uma abordagem com Learning Analytics e séries temporais na análise de dados educacionais**. 2018. Dissertação (Mestrado em Informática Aplicada) – Universidade Federal Rural de Pernambuco, Recife, 2018.

BRASIL, Pedrina Célia. **Um processo analítico de dados educacionais: uma abordagem baseada nos dados socioeconômicos e educacionais dos alunos**. 2019. Dissertação (Mestrado Profissional em Inovação em Tecnologias Educacionais) – Universidade Federal do Rio Grande do Norte, Natal, 2019.

BRITO, Maria Tatiane de Souza. **Um plugin do tipo report para a identificação do risco de evasão na educação superior a distância que usa técnicas de visualização de dados**. 2019. Dissertação (Mestrado em Ciências da Computação) – Universidade Federal da Paraíba, João Pessoa, 2019.

CAMBRUZZI, Wagner Luiz. **GVwise: uma aplicação de learning analytics para a redução da evasão na educação à distância**. 2014. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação Aplicada, UNISINOS, 2014.

COELHO, Orlando Bisacchi. **O papel das representações vetoriais de palavras na pontuação automática de ensaios: uma abordagem baseada em deep learning no contexto de learning analytics**. 2019. 104 f. Tese (doutorado em Engenharia Elétrica e Computação) - Universidade Presbiteriana Mackenzie, São Paulo, 2019.

DAMASCENO, Andre Luiz de Brandão. **Apoiando instrutores na análise de logs dos estudantes de ambientes virtuais de aprendizagem**. 2020. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

DIAS JÚNIOR, Maurício Vieira. **Learning analytics como apoio na avaliação diagnóstica dos docentes no ambiente virtual de aprendizagem**. 2019. Tese (Doutorado em Educação) - Universidade Federal de Alagoas, Centro de Educação, Maceió, 2019.

EDUCAUSE. **EDUCAUSE Horizon Report 2020 - Teaching and Learning Edition**. EDUCAUSE, 2020. Disponível em <https://library.educause.edu/-/media/files/library/2020/3/2020_horizon_report_pdf.pdf?la=en&hash=08A92C17998E8113BCB15DCA7BA1F467F303BA80> acesso em 16 ago. 2023.

FERREIRA, João Luiz Cavalcante. **Md-pread: um modelo para predição de reprovação de aprendizes na educação a distância usando árvore de decisão**. 2016. Dissertação (Mestrado em Computação Aplicada) – Universidade do Vale do Rio dos Sinos, São Leopoldo, 2016.

FERREIRA, Máverick André Dionísio. **Uma abordagem para extração automática de Learning Analytics relacionadas à colaboração em fóruns educacionais**. 2018. Dissertação (Mestrado em Informática Aplicada) – Universidade Federal Rural de Pernambuco, Recife, 2018.

FILATRO, Andrea. **Data Science na educação: presencial, a distância e corporativa**. 1. ed. São Paulo: Saraiva Educação, 2021.

FONTES, Laysa Mabel de Oliveira. **MONITUM: um sistema proativo para monitoramento e avaliação das atividades de tutoria a distância em AVAs**. 2017. Tese (Doutorado em Engenharia Elétrica e de Computação) – Universidade Federal do Rio Grande do Norte, Natal, 2017.

GUEDES, Raphael. **Predição da evasão acadêmica aplicando análise temporal**. 2021. 97 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Goiás, Goiânia, 2021.

HOLANDA, Ana Carla Arruda de. **MOOColab: um framework de colaboração personalizado em Massive Open Online Courses**. 2020. Tese (Doutorado em Ciência da Computação) – Universidade Federal de Pernambuco, Recife, 2020.

LACERDA, Ivan Max Freire de. **Magister - Metodologia de análise de programas de educação à distância baseada em Learning Analytics**. 2018. 41 f. Tese (Doutorado em Engenharia Elétrica e de Computação) - Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2018.

LAGO, Gessyca da Silva. **A interferência da learning analytics na avaliação de impacto e desempenho de bibliotecas universitárias**. 2020. 92 f., il. Dissertação (Mestrado em Ciência da Informação). Universidade de Brasília, Brasília 2020.

LANG, Charles; SIEMENS, George; WISE, Alyssa Friend; GAŠEVIĆ, Dragan; MERCERON, Agathe (coord.). **The Handbook of Learning Analytics – second edition**. SOLAR Society for Learning Analytics Research, 2022. Disponível em < <https://www.solaresearch.org/publications/hla-22/>>. Acesso em 16 ago. 2023.

LUZ, Sandro Fabiano da. **Mensuração da aprendizagem por meio de ferramentas de Learning Analytics no ensino superior**. 2017. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação) – Universidade Federal do Paraná, Curitiba, 2017.

MARTINS FILHO, Vilson. **Design thinking e a criação de ativos do conhecimento na atividade docente**. 2016. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Centro Tecnológico, Florianópolis, 2016.

MARTINS, Robson Douglas da Silva. **Learning Analytics no apoio, planejamento e avaliação de metodologias ativas no Ensino de Física**. 2023. Dissertação (Mestrado em Física Biomolecular) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2023. DOI: 10.11606/D.76.2023.tde-06072023-094213.

MELO, A. S. da C. **Previsão automática de evasão estudantil: um estudo de caso na UFCG**. 2016. 54 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Ciência da Computação, Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande, Paraíba, Brasil, 2016.

MELO, Elvis Medeiros de. **Learning analytics e avaliações online: uma metodologia orientada a ciência de dados em grafos**. Orientador: Ivanovitch Medeiros Dantas da Silva. 2023. 168 f. Tese (Doutorado em Engenharia Elétrica e de Computação) - Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2023.

MENDES, Renê de Ávila. **Aplicação da arquitetura lambda na construção de um ambiente big data educacional para análise de dados**. 2017. 88 f. Dissertação (Engenharia Elétrica) - Universidade Presbiteriana Mackenzie, São Paulo.

MÔNEGO, Leomar Cassol. **Monitoramento da interação e colaboração no ambiente Moodle: um estudo de caso**. 2019. Dissertação (Mestrado em Tecnologias Educacionais em Rede) – Universidade Federal de Santa Maria, Santa Maria, 2019.

MOTA, João Batista. **The web as instrument for university's communication and extension: an user study of the website Espaço do Produtor**. 2012. 132 f. Dissertação (Mestrado em Instituições Sociais e Desenvolvimento; Cultura, Processos Sociais e Conhecimento) – Universidade Federal de Viçosa, Viçosa, 2012.

NUNES, Ronnie Carlos Tavares. **Um modelo de perfil de aluno voltado a aplicações de técnicas de Learning Analytics**. 2019. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2019.

OLIVEIRA, Estêvão Domingos Soares de. **Modelo de diagnóstico de dificuldades de aprendizagem orientado a conceitos**. 2016. Dissertação (Mestrado em Informática) – Universidade Federal da Paraíba, João Pessoa, 2016.

PEREIRA, Natana Lopes. **Boas práticas no uso de ambientes virtuais de ensino e de aprendizagem para mediação da estruturação de trabalhos acadêmicos**. 2019. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Campus Araranguá, Programa de Pós-Graduação em Tecnologias da Informação e Comunicação, Araranguá, 2019.

PERINI, Carlos Antônio de Souza. **Learning Analytics aplicado ao curso online da Língua Brasileira de Sinais**. 2019. Dissertação (Mestrado) - Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM).

PORTAL, Cleber. **Estratégias para minimizar a evasão e potencializar a permanência em EAD a partir de sistema que utiliza mineração de dados educacionais e learning analytics.** 2016. Dissertação (Mestrado em Educação) - Programa de Pós-Graduação em Educação, Universidade do Vale do Rio dos Sinos (UNISINOS).

RIBEIRO, Taffarel Brant. **Tamanho ótimo amostral e análise fatorial e correlacional do desempenho de indivíduos sob a influência de plataformas computacionais de apoio ao ensino.** 2018. Tese (Doutorado em Ciência da Computação) – Universidade Federal de Uberlândia, Uberlândia, 2018.

ROCHA NETO, Tobias Ferreira da. **Orientação acadêmica apoiada por learning analytics e visualização de dados.** 2019. 113f. Dissertação (Mestrado Profissional em Inovação em Tecnologias Educacionais) - Instituto Metrôpole Digital, Universidade Federal do Rio Grande do Norte, Natal, 2019.

SANTOS, Carla Beatriz. **Aprendizagem adaptativa aliada a modelos pedagógicos e tutoria para valorização de conhecimentos prévios.** 2019. Dissertação (Mestrado em Educação e Inovação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2019.

SANTOS, José Roberto dos. **Acompanhamento da aprendizagem pelo estudante por meio de Learning Analytics Dashboards no Ambiente Virtual Moodle.** 2020a. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do ABC, Santo André, 2020.

SANTOS, Mariana Luiza. **Uso de dashboards de learning analytics para promover a autorregulação da aprendizagem em cursos de EAD.** 2020b. Dissertação (Mestrado em Tecnologia Educacional) – Universidade Federal do Espírito Santo, Vitória, 2020.

SCHNEIDER, Thais Fernanda. **Análise do nível de maturidade na adoção de learning analytics em instituições de ensino superior do Brasil.** 2022. Dissertação (Mestrado em Tecnologias da Informação e Comunicação) – Universidade Federal de Santa Catarina, Campus Araranguá, Araranguá, 2022.

SILVA, Deise Garrido. **Avaliação de um curso a distância para cirurgiões-dentistas em atenção à saúde de pacientes com doenças crônicas não transmissíveis: estudo longitudinal prospectivo.** 2019. Tese (Doutorado em Odontopediatria e Ortodontia) – Faculdade de Odontologia, Universidade de São Paulo, São Paulo, 2019.

SOUZA, Rafael Castro de. **Aplicação de Learning Analytics para avaliação do desempenho de tutores a distância.** 2016. 108 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal Rural do Semi-Árido, Mossoró, 2016.

SOUZA, Társis Marinho de. **Uma arquitetura para tomada de decisão pedagógica com dados integrados.** 2017. Tese (Doutorado em Ciência da Computação) – Universidade Federal de Campina Grande, Paraíba, 2017.

TABORDA, Leticia Batista. **Uso de Learning Analytics para aferição de evasão de acadêmicos surdos no curso de Letras Libras da Universidade Federal do Paraná.** Dissertação (Mestrado em Gestão da Informação) – Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas, Programa de Pós-graduação em Gestão da Informação. Curitiba, 2019.

VESCOVI, Paulo Victor da Silva. **Análise preditiva na detecção de evasão de alunos no ensino superior privado brasileiro: abordagem de algoritmos de aprendizado de máquina, com base nas perspectivas acadêmicas, financeiras, geográficas e socioeconômicas.** Dissertação (mestrado profissional

MPGC) – Fundação Getúlio Vargas, Escola de Administração de Empresas de São Paulo, São Paulo, 2020.

VIEL, Caio César. **Objetos de aprendizagem multimídia provenientes da captura ubíqua de apresentações multimodais: produção, interação e análise.** 2013. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de São Carlos, São Carlos, 2013.

ZAPPAROLLI, Luciana Silva. **FAG - Ferramenta de apoio à gestão no ambiente virtual de aprendizagem Moodle utilizando técnicas de Business Intelligence.** 2016. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do ABC, Programa de Pós-Graduação em Ciência da Computação.