



Learning Analytics in Higher Education: a Decade in Systematic Literature Review

¹ A. Salas-Martínez, ORCID: 0000-0002-3839-0978 <zS22000343@estudiantes.uv.mx>

¹ A. Ramírez-Martinell, ORCID: 0000-0003-2370-4994 <albramirez@uv.mx>

² S. Martínez-Ramos, ORCID: 0009-0000-1048-4975 <doc-095@itsperote.edu.mx>

¹ Facultad de Estadística e Informática, Universidad Veracruzana,
Xalapa, México.

² Ingeniería Informática, Instituto Tecnológico Nacional de México,
Perote, México.

Abstract. In the last decade, Learning Analytics (LA) has evolved in a positive way, considering that the term emerged in 2011 through the Society for Learning Analytics Research (SoLAR). This area of data analytics can be identified as a specialization of Educational Data Mining (EDM). LA emphasizes student learning outcomes. In addition to, a better understanding of student learning behavior and processes. While EDM focuses on helping teachers and students with the analysis of the learning process using popular data mining methods. The purpose of this research is to explore the first decade of work with the application of Learning Analytics in Higher Education Institutions (HEI) in the context of Tutoring Information Systems (TIS), with the intention of supporting institutions, teachers and students to decrease dropout rates. This article presents a systematic literature review (SLR) with 17 primary studies, comprised between 2014 and 2024. The findings reflect the use of LA in improving or optimizing learning using student academic history obtained through Learning Management Systems (LMS), noting the scarcity of works with a focus on tutoring or academic advising. Ultimately, a gap is opened to apply LA in HEI, with information from Institutional Tutoring Program (PIT), integrated with information from an LMS, to contribute to student permanence.

Keywords: learning analytics; tutoring; counseling; systematic literature review; higher education institutions.

For citation: Salas-Martínez A., Ramirez-Martinell A., Martínez-Ramos S. Learning analytics in higher education: a decade in systematic literature review. Trudy ISP RAN/Proc. ISP RAS, vol. 36, issue 6, 2024. pp. 215-230. DOI: 10.15514/ISPRAS-2024-36(6)-12.

Acknowledgements. This research was funded by the Universidad Veracruzana through the Faculty of Statistics and Informatics, the coordination of the PhD in Computer Science and the National Council for the Humanities, Sciences and Technologies (CONACYT) through grant number 730418. We also thank the Instituto Tecnológico Superior de Perote for their support in the development of the research.

Учебная аналитика в высшей школе: десятилетие в систематическом обзоре литературы

¹ А. Салас-Мартинес, ORCID: 0000-0002-3839-0978 <zS22000343@estudiantes.uv.mx>

¹ А. Рамирес-Мартинель, ORCID: 0000-0003-2370-4994 <albramirez@uv.mx>

² С. Мартинес-Рамос, ORCID: 0009-0000-1048-4975 <doc-095@itsperote.edu.mx>

¹ Университет Веракрус, факультет статистики и информатики,
Халапа, Мексика.

² Национальный технологический институт Мексики, Компьютерная инженерия,
Пероте, Мексика.

Аннотация. Термин учебная аналитика (LA) был введен Обществом исследований в области аналитики обучения (SoLAR) в 2011 году, и в последнее десятилетие это направление развивалось самым позитивным образом. Эта область аналитики данных может быть определена как интеллектуальный анализ образовательных данных (EDM). В дополнение к тому, что LA позволяет лучше понять поведение и процессы обучения студентов, учебная аналитика также обеспечивает более четкое выявление результатов обучения студентов, в то время как EDM фокусируется на оказании помощи преподавателям и учащимся при анализе процесса обучения с использованием популярных методов интеллектуального анализа данных. Целью данного исследования является изучение первого десятилетия работы с применением методов учебной аналитики в высших учебных заведениях (ВЕИ) в контексте обучающих информационных систем с целью поддержки учреждений, преподавателей и студентов для снижения показателей отсева. В этой статье представлен систематический обзор литературы с 17-ю первичными исследованиями, проведенными в период с 2014 по 2024 год. Полученные результаты отражают использование LA для улучшения или оптимизации обучения с использованием академической истории студентов, полученной с помощью систем управления обучением, и отмечают нехватку работ с акцентом на обучение или академическое консультирование. В конечном счете, вскрыта недостаточность применения методов LA в ВУЗах к информации от Программы институционального обучения, интегрированной с информацией от систем управления обучением, которые могли бы внести свой вклад в постоянство студентов.

Ключевые слова: учебная аналитика; преподавание; консультирование; систематический обзор литературы; высшие учебные заведения.

Для цитирования: Салас-Мартинес А., Рамирес-Мартинель А., Мартинес-Рамос С. Учебная аналитика в высшей школе: десятилетие в систематическом обзоре литературы. Труды ИСП РАН, том 36, вып. 6, 2024 г., стр. 215–230 (на английском языке). DOI: 10.15514/ISPRAS–2024–36(6)–12.

Благодарности. Грантом № 730418 исследование финансировалось Университетом Веракрус через факультет статистики и информатики, аспирантуру в области компьютерных наук, а также Национальный совет по гуманитарным наукам, наукам и технологиям (CONACYT). Мы также благодарим Высший технологический институт Пероте за поддержку исследования.

1. Introduction

The use of *learning analytics* (LA) in the online learning environment has increased exponentially, because its application can help institutions, teachers and tutors with problems such as decision making and measurement of student success, considering the digital footprint that can be obtained from students in each *Higher Education Institution* (HEI). Currently it is a reality to mention that Higher Education has been forced to retake the use of technological tools such as *Learning Management Systems* (LMS) due to the arrival of the COVID-19 pandemic in two thousand twenty, because although these tools are not new, there was a resistance to their use in an official way. However, LMS gained great importance after the pandemic, tools such as Google Classroom, Microsoft Teams and Moodle are some of the most implemented by HEI in the world [1].

When talking about LA, it is important to mention that it is an incipient term, which arises as a specialization of Educational Data Mining (EDM), which first appeared in 2005 at the first workshop

on Educational Data Mining [2], being in 2008 the 1st International Conference on Educational Data Mining, held in Montreal, Quebec. While LA arises in the summer of 2012, being in the *Society for Learning Analytics Research (SoLAR)* where they define it as: “*the measurement, collection, analysis and reporting of data about learners and their contexts, in order to understand and optimize learning and the environments in which it occurs*” [3]. Meanwhile, EDM is defined as “*An emerging discipline concerned with developing methods for exploring unique types of data that come from educational settings and using those methods to better understand the learner and the environments in which they learn*” [2]. Therefore, we can say that first, LA places greater emphasis on student learning outcomes, a better understanding of student learning behavior and processes, in addition to better educational research. Second EDM focuses on helping teachers and students with the analysis of the learning process using popular data mining methods.

Some of the areas that have been considered for applying these analytics are:

1. Create alerts for stakeholders: monitor students’ academic progress to quickly and correctly identify negative student behaviors, such as lack of motivation, dropping out, etc;
2. Group/profile students: separate students into groups according to their individual characteristics, personality traits, preferred learning methods, and other considerations;
3. Predict student performance: by calculating a performance estimate of grades, knowledge or score.

However, it has been observed that there has been little exploration of the academic tutoring that HEI students receive.

It is worth emphasizing that tutoring is an institutional program that arises with the intention of supporting students in their academic, personal and professional processes during their education in any HEI in Mexico. Therefore, Institutions such as Normal Schools, Universities and Technological Institutions, have their respective tutoring programs dedicated to support students during their academic life. Thus, from these systems it is possible to identify diverse situations of each student, from economic, health, academic and social points of view, problems that can directly impact their academic performance and even cause failure and in the worst-case scenario, desertion [1]. Therefore, this article is an extended version of the research published by Salas et al. 2024 [1], where the updates of the last decade to perform SLR, the discussion achieved during the presentation of the same at the *11th International Conference on Research and Innovation in Software Engineering (CONISOFT 2023)*, as well as the integration of new findings, are considered. The integration of new findings, are part of the content that can be found in this publication. It aims to provide the reader with the following points:

1. Tutoring Information Systems (TIS) that use LA;
2. Identify on whom these TIS are focused on;
3. What are the most common interests in TISs;
4. How can SITs be quantified and categorized;
5. How the IES use the LA for tutoring;
6. How do they interpret and visualize LA-based data in the IES;
7. What information has been analyzed in the SIT of the IES and what they used for the analysis.

2. Background and related work

For contextualization it is important to define that according to Siemens [4], LA is defined as “*The measurement, collection, analysis and reporting of data about learners and their contexts, in order to understand and optimize learning and the environments in which it occurs*”. With the implementation of LA it is possible to find out more hidden information about learners in online learning. For this reason, it plays a relevant role in online learning whose main interest is to identify problems with learning and

improve the learning environment.

The following are some of the papers that were reviewed prior to conducting SLR with the intention of getting into the context of the work on Learning Analytics at HEI. In a study conducted by a university in Korea, empirical validation of the effects of a *Learning Analytics Dashboard (LAD)* was sought the results of this study, it was observed that students who interacted with the LAD scored higher compared to those who did not use it [5]. These results mark a path with respect to the need to review the LAD with features that motivate and support students who have different levels of academic performance [5].

On the other hand, a second study proposes the question: How do we begin the institutional adoption of learning analytics? This question is a common one among faculty, administrators, and researchers who seek to conduct Learning Analytics (LA) [1]. In summary, this study builds on established models for the adoption of business analytics, showcases two projects conducted in Australia, to develop and evaluate approaches for LA adoption in HEI [3]. The focus of the study highlights the importance of the socio-technical nature of LA and the complexities relevant to adoption in HEI.

A third study aimed to investigate student expectations regarding the characteristics of learning analytics systems and the willingness to use these features in learning. It was an exploratory and qualitative study, applied to 20 university students, who were interviewed about their expectations about learning analytics features. The findings of the study were complemented with a quantitative study applied to 216 students [6]. As results it was found that students expect Learning Analytics functions to support their planning and organization of learning processes, as well as provide self-assessments, adaptive recommendations, and produce personalized analyses of their learning activities [7].

Among the studies conducted, there is also "*Learning Dashboard for Insights and Support during Study Advice (LISSA)*", a LAD designed, developed and evaluated in collaboration with advisors, which aims to facilitate communication between advisors and students through the visualization of qualifications that are available in the HEI. The study found that the dashboard supports the ongoing dialogue between the advisor and student, motivating students, activating the conversation and providing tools for personalization, depth and nuance to the advising session, providing information at the factual, imperative and reflective levels, and engaging those involved in an active role during the session [7].

In another study, the use of a *Learning Analytics Dashboard (LAD)* to inform the teaching of five university professors was investigated using qualitative inductive analysis to identify salient emergent themes. The results of the study showed that instructors did not always draw on analytics with specific questions, but rather with general areas of curiosity [1]. The findings were synthesized into an analytical model of instructor use that provides useful categories of activities for future study and support [3].

On the other hand, an empirical study was identified, whose objective was to analyze both intrinsic and extrinsic motivation factors perceived through LMS such as Moodle. Said study is based on the *self-determination theory (SDT)*, the findings of the study reveal that intrinsic and extrinsic motivation significantly influence the effectiveness of student-perceived learning and the improvement of academic performance [8].

Another of the studies reviewed, focuses on determining the motivation of students in LA context, in this study aims to perceive the state of student motivation at a high level of abstraction. The results showed that it is possible to perceive the state of student motivation at a high level of abstraction [9]. Finally, one study identified was student-oriented, providing information and promoting self-regulated learning. In this study, a LAD design aligned with SRL (*Self-Regulated Learning*) theory was created, which was called *My Learning Analytics (MyLA)* [10], which seeks to better understand how students use a learning analytics tool. The study consisted of performing a sequential analysis of student interactions with three different dashboard visualizations implemented in an LMS. The results of this study showed discriminatory patterns in the use of the dashboard between different levels of

academic performance and self-regulated learning, particularly for students with low performance and high levels of self-regulation. The finding of this study highlights the importance of differences in students' experience with a student-oriented dashboard and emphasizes that one type of dashboard does not fit all in the design of learning analytics tools [11].

3. Research method

The research process was initiated through a Systematic Literature Review (SLR) in order to use explicit and systematic procedures as opposed to traditional research. Therefore, it followed Kitchenham and Charters' [12] guidelines on SLR in software engineering and Zhang's guidelines [13] proposed under the concept of '*Quasi-Gold Standard (QGS)*' applied in the identification of relevant software engineering studies. It is worth mentioning that Kitchenham's methodology is based on three phases that are planning, conducting and documentation, to achieve the identification of relevant studies. However, in this research has been combined with the methodology proposed by Zhang, to provide greater rigor in the search for relevant studies, considering that Zhang contemplates the automatic search and manual search to identify relevant studies, in addition to applying a sensitivity and accuracy assessment that confirm the rigor of the SLR. In addition to considering some recommendations from the works [14-15] the review proposed in this paper is composed of two subsections: Planning and Conduction.

3.1 Planning

In this phase, the formulation of the research questions was carried out, the search process was established, and its description follows.

3.1.1 Research questions

To drive the review process, the following seven research questions were generated, where each question seeks to clarify the panorama on the application of Learning Analytics in Higher Education Institutions (HEI).

1. **[RQ1]** Are there Tutoring Information Systems (TIS) used by the HEI where LA is used?
2. **[RQ2]** In the approaches used, is was at the center the student at the center, teaching tasks or tutorial management?
3. **[RQ3]** What are the declared interests identified in the TIS in HEI?
4. **[RQ4]** How can the TIS used in HEI be quantified and categorized?
5. **[RQ5]** How does HEI use LA for TIS?
6. **[RQ6]** How does the TIS use in HEI interpret and visualize LA-based data?
7. **[RQ7]** What information has been analyzed from the TIS and what was used to analyze the information?

3.1.2 Search process

The search process in this article followed the "*Quasigold standard*" (QGS) strategy [13]. This process consists of five steps: 1) Identify related databases, 2) Establish the QGS, 3) Define or obtain the search string, 4) Perform the automatic search and 5) Evaluate the performance of the search. Each of the steps is described below in the context of the investigation.

1) Identify related databases

In this phase, journals were selected for the manual search and databases (DB) digital libraries and indexing services for the automatic search to. The following six journals were considered: *Knowledge and Learning, Technology, Informatics in Human Behavior*, for the SLR decade update the following journals were added: *Information Development, International*

Journal of Instruction and Perspectives on Psychological Science, for their relevance to the topic to be addressed, as well as for the Educational Institutions participating in the edition, the impact factor they maintain and the periodicity of the journal. Six others, by automatic search, coverage, overlap and accessibility of libraries and search engines. The following were included: *IEEE Xplore*, *ACM Digital Library*, *Springer Link*, *ScienceDirect*, *Wiley Online Library* and *EBSCOhost Academic Search*; the selected databases are available in the information resources of the *Consortio Nacional de Recursos de Información Científica y Tecnológica (CONRICyT)* provided by the Universidad.

2) **Establish the QGS**

In this phase, the inclusion and exclusion criteria are defined, after which a manual search is carried out in the previously selected journals, which consists of analyzing all the volumes and identifying the articles that meet the established criteria, in this update of the SLR the last decade is contemplated, from 2014 to February 2024 and the criteria of the first version of the SLR are maintained; Table 1 shows the criteria established in this Systematic Literature Review (SLR).

Table 1. Inclusion And Exclusion Criteria.

Inclusion		Exclusion	
Id	Description	Id	Description
IC1	Access to the publication is through National Consortium for Scientific and Technological Information Resources (CONRICyT) provided by the university.	EC1	The publication is an exact duplicate of a study obtained from another search engine
IC2	The publication date is from 2014 to February 2024	EC2	The publication is not in Spanish or English
IC3	The publication must be a research article on software, learning analytics and tutoring in higher education institutions (Journal Article)	EC3	The full text is restricted in the retrieved publications
IC4	The publication must reference at least two search terms	EC4	The publication is not applied in Higher Education Institutions
IC5	The publication must answer at least one research question		

3) **Define or obtain the search string**

At this point in the review, the terms “*Learning Analytics*”, “*Tutoring System*”, “*Academic advising*”, “*Academic counseling*” and “*Higher Education Institutions*” were taken as reference. It should be mentioned that depending on the databases (DB) consulted, the search string was refined and adapted depending on the fields available in the advanced search of each database (DB) [1]. Table 2 shows the search string used in a general way in all the previously mentioned search engines.

Table 2. Search String Executed.

Search String
(“Learning Analytics”) AND (“Mentoring system” OR “Academic Advising” OR “Tutoring System” OR “Academic Counseling”)

4) **Perform the automatic search**

In this phase, the search was performed in each of the database (DB) selected by applying the specific syntax in each of them. In this update, a total of 157 publications were obtained, achieving an increase of 86 studies linked to our search string, it notes, unlike the SLR presented previously [1], the 157 studies found were verified by applying the inclusion and

exclusion criteria to obtain the final corpus.

Selection of Primary Studies. To carry out the selection of these studies, it was necessary to apply the process of inclusion and exclusion criteria, which consisted of three stages, as shown in Table 3. This organization served to reduce the number of publications while retaining the relevant studies for subsequent analysis.

Table 3. *Study Selection Process.*

Stages	Criteria
Stage 1	IC1, IC2 and EC1
Stage 2	IC3, IC4, EC2 and EC3
Stage 3	IC5 and EC4

The following is the selection of candidate studies after application of the inclusion and exclusion criteria, as shown in Table 4.

Table 4. *Results of the selection of candidate studies.*

Id	Source	Candidate papers	Filter 1	Criteria I/E	Included
1	IEEE	21	0	17	4
2	ACM	22	4	15	3
3	Springer	28	0	16	12
4	ScienceDirect	82	67	7	9
5	Wiley OnLine	2	0	0	2
6	EBSCOhost	2	0	1	1
Total		157	71	56	30

5) Evaluate the performance of the search

At this point the results of the automatic search are compared with the manual search (QGS). To achieve this, we used the equations proposed by Zhang et al. [13]. First, the Equation sensitivity or recovery was calculated Eq. (1), to obtain the number of relevant studies retrieved, we subtract from the total number of studies retrieved automatically, which are 157. 140 studies were not relevant. To obtain the total number of relevant articles, we divided it by the total number of Relevant studies found, thus achieving 100% of the corpus $[(157-140/17)*100]$. Afterwards, to calculate the precision we found 17 studies found by QGS. 10 are not relevant, therefore, we proceeded to use the Eq. (2), to obtain the precision, the number of relevant studies recovered, subtracting the 71 studies obtained through the automatic search, the 10 studies that are not relevant and then we divide it by the total number of studies retrieved in the automatic search, thus obtaining 93% of the corpus $[(157-10/157)*100]$. Therefore, it is identified that both parameters are within the suggested threshold that indicates the percentages must be greater than 70% to be acceptable, the maximum Equation sensitivity can be observed in Eq. (3) and the optimal precision in Eq. (4).

$$Sensitivity = \frac{NRSR}{TNRS} 100\% \quad (1)$$

$$Precision = \frac{NRSR}{NRS} 100\% \quad (2)$$

where: *NRSR* = Number of relevant studies retrieved
TNRS = Total number of relevant studies
NSR = Number of studies retrieved

$$Sensitivity = \frac{157 - 140}{17} 100\% = 1.0 \quad (3)$$

$$Precision = \frac{157 - 10}{157} \cdot 100\% = 0.93 \tag{4}$$

3.2 Conduction

3.2.1 Quality assessment

At this point in of the process, the 30 selected studies are taken up again and again subjected to validation to identify only those studies that meet the necessary quality; therefore, the quality assessment instrument was prepared, which contains the quality control questions. A value of 1 was assigned to the questions that are answered with the word yes, a value of 0.5 for those questions that are considered to be partially compliant and 0 for those that are not. Table 5 shows the questions asked to assess the quality of the study.

All the studies found were evaluated with the proposed instrument to guarantee the quality of the chosen studies. The possible score to achieve was between 0 and 8 points. After the evaluation, all the studies that achieved a score greater or equal to 6.5 were considered; see Table 6. It was observed that 56% (17 studies). See Table 7 met the established quality criteria for the most part, while 43% (13 studies) did not meet them, therefore they were discarded from the final selection.

3.2.2 Data Extraction

This phase consisted of extracting the most relevant data from each of the primary studies identified, with the support of the *Parsifal* platform. Bibliographic information was extracted for each study such as: title, authors, year of publication, source, type of publication, DOI, keywords and abstract. It also includes information that helps to answer the research questions.

4. Results

This section of the article presents different results obtained during the SLR of the research, such as: text analysis applied to the bibliography of the corpus through the *VosViewer software*, frequent word cloud of the content of the articles generated from *MAXQDA 2022*, in addition to answering the research questions. A narrative synthesis is provided based on the data identified in the research corpus. To start, an overlay visualization generated by *VosViewer* is shown, see Fig. 1. It is possible to observe the grouping through four clusters formed with titles and summary of the corpus, where the terms *Learning Analytics*, *Educational data Mining*, *Environment* and *Dashboard* are found, demonstrating the linkage and relevance they present in the research. On the other hand, as shown in Fig. 2, relevant works within the SLR are scarce, however, the application of LA in HEI is a topic of interest from 2015 to date, observing a positive trend of articles towards 2024.

4.1 Answers to research questions

[RQ1] Are there TIS used by HEI where LA are used? Based in the work carried out by Chatti et al. [32] who proposes a reference model for Learning Analytics (LA) based on four specific dimensions, (1) *what* (e.g., data, environment and context), (2) *why* (e.g., objectives), (3) *how* (e.g., techniques/methods) and (4) *who* (e.g., stakeholders), model that helps to have an overview of LA and its concepts of relevance, in addition to including the review carried out by Bodily et al. [17] where they categorize the works under an approach between various subfields of educational technologies. In this review we analyze the objectives and technologies that guide those interested in making effective decisions about teaching, within the analysis we can identify the following categories to classify the most relevant jobs in the educational field and LA, see Table 8.

[RQ2] In the approaches used, is the student at the center, the teaching tasks or the management of tutorials? Based on the studies reviewed, it can be classified that the works where Learning Analytics (LA) is applied are mainly focused on the following actors [16]: teachers, students, tutors and researchers, with the latter having less presence in the research reviewed [25, 28]. According

to Robert Bodily [17] in his review of student-oriented learning analysis dashboards and educational recommender systems, most of the systems found are oriented 74 percent to the instructor, and he also states that researchers do not conduct much research on the impact of the systems on teaching and learning. Also, Perez Sánchez [27], Ranjeeth [23] and Rafique [26] focus on students.

Table 5. Quality Assessment Instrument Used for Included Study Evaluation.

Id	Question
QA01	Are the objectives, research questions, and hypotheses (if any) clear and relevant?
QA02	Is there an adequate description of the context in which the research was conducted?
QA03	Is the suitability of the case to address the research questions clearly motivated?
QA04	Are the case and its units of analysis well defined?
QA05	Is the case study based on theory or linked to existing literature?
QA06	Are the data collection procedures sufficient for the purpose of the case study (data sources, collection, validation)?
QA07	Are ethical issues (personal intentions, integrity, confidentiality, consent, review board approval) adequately addressed?
QA08	Is a clear chain of evidence established from observations to conclusions?

Table 6. Primary Studies Quality Assessment.

Id	QA1	QA2	QA3	QA4	QA5	QA6	QA7	QA8	Total
PS01	1	1	1	1	0.5	0.5	1	1	7.0
PS02	1	0.5	1	0.5	1	1	1	1	7.0
PS03	0.5	1	1	1	1	1	1	1	7.5
PS04	1	1	1	1	1	1	1	0.5	7.5
PS05	1	1	0.5	1	0.5	1	1	1	7
PS06	1	1	1	1	1	1	1	1	8.0
PS07	1	1	1	1	1	0.5	0.5	0.5	6.5
PS08	1	1	1	1	1	1	1	1	8
PS09	1	1	1	0.5	1	1	0.5	1	7.0
PS10	1	1	1	1	1	1	0.5	1	7.5
PS11	1	1	1	1	1	0.5	0.5	0.5	6.5
PS12	1	1	1	1	1	0.5	0.5	1	7.0
PS13	1	1	1	1	1	1	0.5	0.5	7.0
PS14	1	1	1	0.5	1	1	1	0.5	7.0
PS15	1	1	1	0.5	0.5	0.5	1	1	6.5
PS16	1	1	1	0.5	1	0.5	0.5	1	6.5
PS17	1	1	1	0.5	1	1	0.5	1	7.0

Table 7. Primary Studies.

Id	Author	Year	Database
PS01	Reyes et al. [16]	2015	Springer
PS02	Siemens et al. [4]	2015	ACM
PS03	Bodily et al. [17]	2017	IEEE
PS04	Rojas et al. [18]	2017	EBSCOhost
PS05	Tempelaar et al. [19]	2017	IEEE
PS06	Viberg et al. [20]	2018	ScienceDirect
PS07	Herodotou et al. [21]	2019	Springer
PS08	Guerra et al. [22]	2020	Wiley
PS09	Ranjeeth et al. [23]	2020	ScienceDirect
PS10	De Laet et al [24]	2020	Wiley
PS11	Guzman et al. [25]	2021	Springer
PS12	Rafique et al. [26]	2021	IEEE

PS13	Perez et al. [27]	2022	Springer
PS14	Kaliisa et al. [28]	2022	ScienceDirect
PS15	Hao et al. [29]	2023	ScienceDirect
PS16	Prinsloo et al. [30]	2023	Springer
PS17	Kaur et al. [31]	2023	IEEE

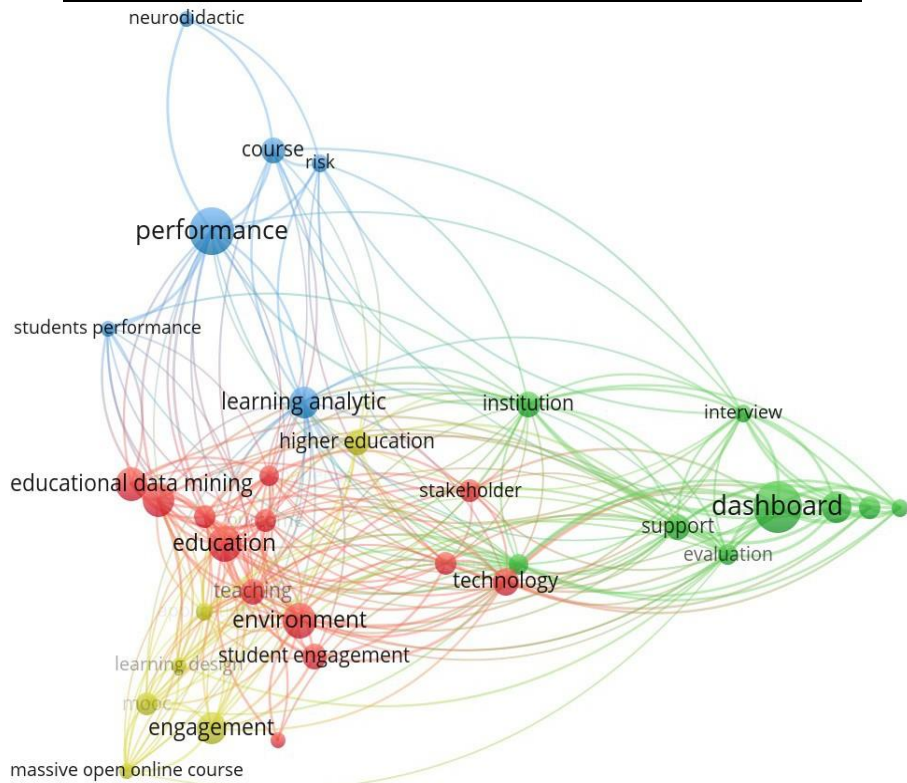


Fig. 1. An overlay visualization from research corpus.

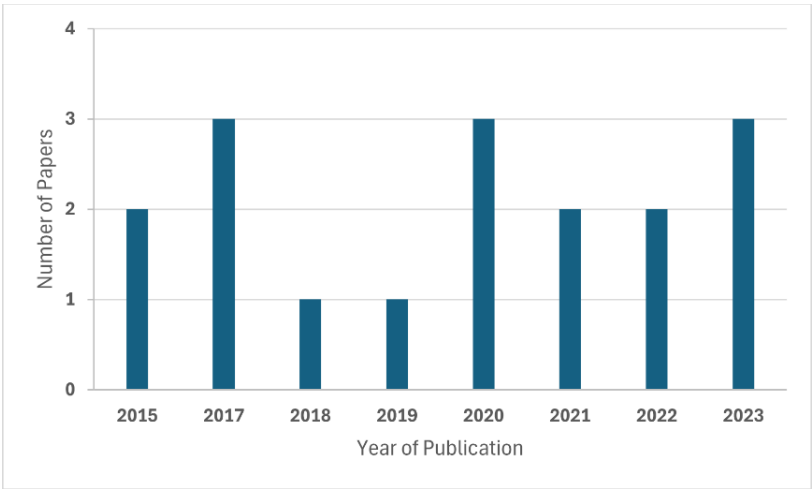


Fig. 2. Year of publication.

Table 8. Categories of systems using learning analytics in higher education institutions.

Category	Article
Intelligent tutorial systems	[17]
Predictive systems	[21], [26]
Academic performance system	[17],[20]
Educational recommendation systems	[27], [26]
Learning boards	[17], [22], [21], [26], [28], [20]
Educational data mining system	[17], [27]

For example, in Ranjeeth's literature study, some of the predictions that have been made are: predictions of student grade point average (GPA), prediction of student performance in graduate programs, prediction of instructor performance and likely student performance in gaining admission to college, prediction of attrition from college programs, and prediction of student grades using social network theory analysis. For his part Rafique says that student performance can be predicted from the student's digital fingerprints [22], i.e., demographics, behavior, facial emotion control records while using an intelligent tutoring system [21]. While for researchers the most predominant focus is on an online learning environment to predict student performance and timely intervention, however, Rafique expresses that it is very limited work in traditional learning environments.

[RQ3] What are the declared interests that are identified in the TIS in HEI? According to the research work conducted by Bodily [17] in 2017, it is possible to identify that there is an interest in identifying student-oriented LA reporting systems with respect to their purpose, functionality and the types of data collected. Also, Schwendimann et al. [17] mentions the interest in the mechanisms by which student-oriented systems attempt to improve teaching and learning, which requires analysis through different categories such as type of data, target users, and evaluation. Learning Analytics Dashboards (LAD) also identified by their acronym LAD were found to have evaluated categories such as: goal orientation, usefulness of information, visual effectiveness, ease of use, comprehension, reflection, motivation for learning, behavior change, performance improvement, and competency development.

[RQ4] How can the TIS used in HEI be quantified and categorized? In the various works reviewed, multiple approaches and objectives have been observed, however, none of them gives an answer to this question since most of the cases where tutoring is discussed, their focus is on intelligent tutoring systems, as identified by Bodily, to this point the only ones that come closest to working with tutoring in Latin America, are the systems generated by Learning Analytics in Latin America (LALA) expressed in the research of Guerra et al. [22]. This arises from a framework called COALA [24] (Context Adaptation for Learning Analytics), which is constituted by four dimensions for adapting tools: objectives of using a Learning Analytics Dashboards (LAD) (e.g., identifying subjects in which students have low or high performance), stakeholders (e.g., advisors, teachers, students and administrative staff), key moments in which the use occurs (e.g., at the beginning of the academic year, when a course is registered or when they receive grades) and the interaction of stakeholders (e.g., face-to-face sessions with the advisor-student). This project was conducted within the context of Latin America partner institutions, the participating institutions were *University of Cuenca in Ecuador* (Cuenca), *University Austral of Chile* (UACH) and *Polytechnic Superior School of the Litoral in Ecuador* (ESPOL) [24].

[RQ5] How do HEI use LA for TIS? Among the works reviewed, there are three cases where the use of LA applied to tutorial information systems has been most closely approached. Identified in the following institutions: *University of Cuenca*, *University Austral of Chile* and *Polytechnic Superior School of the Litoral*. See Table 9. The three cases coincide in combining information from the curricular structure and academic records to observe student progress [1]. However, the three Latin American universities adapted an advisory board, originally implemented at KU Leuven in Belgium. In all three cases, the

context was the main factor for adapting the dashboard, taking up that the LALA project [22] focuses on four different elements of the context such as: Objective, Actors, Key Moments and Interactions.

[RQ6] How do TIS used in HEI interpret and visualize the data based on LA? Several institutions have begun to adopt Predictive Learning Analytics (PLA) [21], they use a number of computational techniques (e.g., Bayesian modeling, cluster analysis, predictive modeling) to identify which students will pass a course and which are at risk. According to Merceron’s categorization [20], it is identified that predictive methods (regression and classification) with 32% are considered the most frequent, below are relationship mining methods (association rules, correlations, sequential patterns and causal data mining) and methods for distilling data for human judgment tied with 24% frequency, where statistics and visualization are included. According to Viberg et al. [20] the application of methods for data analysis has been increasing from 2014 to 2017, reflecting from 2017 an increase in mining methods compared to previous years. In the work of Kim et al. [34], k-medoids clustering and Random Forest classification followed by logistic regression were applied for the analysis of the identified cluster profiles to analyze students’ self-learning patterns in asynchronous mode. Likewise, to predict whether students pass or fail the course, the following models were used: Random forest (RF), K-Nearest Neighbors (KNN), logistic regression (LR), neural networks (NNETS), TreeBagging (TB) and Bayesian additive regression trees (BART) Perez et al. [27].

Table 9. Some Applications of Learning Analytics in Higher Education Institutions.

Classification	Name	Goal	Author	Country
Learning Analytics Dashboard (LAD)	OUA dashboard (Open University Analyze)	Helps teachers identify at-risk students in online course	Herodotou et al. [21]	United Kingdom
	ESPOL LAD (Escuela Superior Politecnica del Litoral)	Support student-advisor dialogue when advising study plan in student	Guerra et al. [22]	Belgium
	LISSA (Learning dashboard for Insights and Support during Study Advice)	Support student–advisor dialogue focus on first-year students	Charleer et al. [7]	Belgium
	LADA (Learning Analytics Dashboard for Advisors)	Support advice on study plan by advisors	De Laet et al. [24]	Belgium
	AvAc (Advising dashboard Avance Academico)		Gutierrez et al. [33]	Cuenca
	TrAC dashboard (Trayectoria Academica y Curricular)	Inspired by dashboard LISSA superimposes the academic records on the curricular structure, since this is the “natural” way in which academic progress is understood in the institution	Gutierrez et al. [33]	Ecuador
Framework	COALA framework (Context Adaptation for Learning Analytics)	To evaluate the support provided by the adapted dashboard comprises three modules including a visualization module, a module for group formation and intervention, and a prediction module	Guerra et al. [22]	Chile
	Smart Learning		Rafique et al. [26]	Pakistan
	SHEILA (Supporting Higher Education to Integrate LA)	The proposed framework will enhance systematic adoption of learning analytics on a wide scale	Viberg et al. [20]	United Kingdom

[RQ7] What information has been analyzed from the TIS and what did they use for the analysis of the information? It has been identified that to date only the data available in some of the educational platforms have been used to generate conclusions from the LA perspective. Among the data identified we can find: student demographics: age, gender, disability, previous grades, ethnicity, successful completion of previous courses, previous experience of the student at the university (new versus continuing student), best score in the previous course and sum of credits earned [21]. In the work of Perez et al. [27], characteristics such as: LMS, numbers of accesses, participation scores, learning activity ratings, submissions, published content, completed learning activities and peer reviews were contemplated, where these characteristics presented significant statistics between students who failed and those who passed.

5. Discussion

With the systematic review of literature, we have realized that there is little research that considers the tutoring received by students in higher education institutions, through student support programs, such as the Institutional Tutoring Program (PIT), which is promoted by institutions such as the UN (United Nations) and ANUIES (National Association of Universities and Higher Education Institutions) in Mexico. At the same time, it has been observed that when researchers make use of data, these are only data extracted from LMS, or application of surveys. Therefore, an important gap opens up for applying LA by integrating data ecosystems that can be made up of data extracted directly from the LMS database, data that are generated manually by tutoring coordinators in HEIs, and even surveys applied for concrete measurements such as emotion, engagement, and motivation, to mention a few. It is worth stressing that working with this type of data is a major challenge, as it implies good data quality and scope, as well as privacy and ethics in working with the data. On the other hand, it is also clear that most of the works that present advances in LA issues are from the USA and European countries. Thus, it is necessary to bet on the application of LA in HEIs, to contribute to the institutions by supporting the permanence of students, to teachers, letting them know what is happening in the student's learning process, and to the latter, providing recommendations for them to enjoy their learning process in the best possible way. The challenges faced by learning analytics focused on student behavior are the integration of data sets from diverse environments, advances in technology, and ethical problem solving. Despite these challenges, this research aims at solutions that contribute early on by making recommendations or warnings to achieve student retention.

6. Conclusion

With the SLR on the subject it has been identified that the term LA as we describe it today arises from the year 2011, being the most cited definition the one that arises in the "*1st International Conference on Learning Analytics and Knowledge 2011*". Therefore, we can say that it is clear at this time to identify that the objective of LA is to improve learning. It should be noted that, being considered a new term, it is often difficult for some to identify the differences with its predecessor, the term Educational Data Mining. It should be noted that although both work with educational data, their approaches are different. Firstly, we have EDM which focuses on helping teachers and students with the analysis of the learning process using popular data mining techniques such as: clustering, association and classification to name a few, we could say that it seeks the transformation of data into relevant information. Unlike LA where the emphasis is on learning outcomes, a better understanding of the learner's behavior and processes, therefore, its objectives are recommendations, predictions, adaptation and personalization, in addition some common methods are usually used such as: classification, clustering and association, in addition to, social network analysis, sentiment analysis, prediction of learner success among others. It should also be noted that the term Tutoring in this context is defined as a process of group or individual accompaniment that a tutor provides to a student during his stay in an IES, with the purpose of contributing to his integral formation, besides influencing the fulfillment of the institutional goals related to the educational quality such as: raising

the terminal efficiency rates and decreasing the failure and desertion rates. On the other hand, in order for LA to achieve this objective, different techniques and methods are used, which are applied to the data offered by the educational platforms. It is important to mention that there are still few studies on this topic, but it has been identified that in Latin America LA has already begun to be used in HEIs. As indicated, it is still a little explored topic with great areas of opportunity. All the studies found present the common denominator of the use of student information regarding their academic history or surveys as instruments for specific measurements such as academic performance, emotions or student motivation during the course; however, there are still minimal cases in which other types of information are used to predict their future behavior. This research opens a gap to resume studies that can integrate academic information that can come from the SIS (Student Information System), LMS tracking data and information from institutional tutoring programs. It should be noted that the complexity of this part will depend on the ease of access to data from these tutoring programs, since there is no specific system or standard for integrating the information, so each institution stores and processes its data in a particular way. The challenge we face with LA is the integration of data ecosystems from different sources, as well as data quality and scope, and data privacy and ethics.

References

- [1]. A. Salas Martinez and A. Ramirez Martinell, Application of learning analytics in higher education institutions: A systematic literature review, in 11th International Conference in Software Engineering Research and Innovation (CONISOFT) (2023).
- [2]. G. Siemens and R. S. J. d. Baker, Learning analytics and educational data mining: towards communication and collaboration, in Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (2012), p. 252–254.
- [3]. D. Gasevic, Y. Tsai, S. Dawson, and A. Pardo, How do we start? an approach to learning analytics adoption in higher education, *International Journal of Information and Learning Technology* 4, 342 (2019).
- [4]. G. Siemens and P. Long, Penetrating the fog: Analytics in learning and education, *EDUCAUSE Review* 5, 30 (2011).
- [5]. S. J. Aguilar, S. A. Karabenick, S. D. Teasley, and C. Baek, Associations between learning analytics dashboard exposure and motivation and self-regulated learning, *Computers & Education* 162 (2021).
- [6]. C. Schumacher and D. Ifenthaler, Features students really expect from learning analytics, *Computers in Human Behavior*, 397 (2018).
- [7]. S. Charleer, A. Moere, J. Klerkx, K. Verbert, and T. D. Laet, Learning analytics dashboards to support adviser-student dialogue, *IEEE Transactions on Learning Technologies* 11, 389 (2018).
- [8]. M. Waheed, K. Kaur, N. Ain, and N. Hussain, Perceived learning outcomes from moodle: An empirical study of intrinsic and extrinsic motivating factors, *Information Development* 32, 1001 (2016).
- [9]. O. Talbi and A. Ouared, Goal-oriented student motivation in learning analytics: How can a requirements-driven approach help? *Education and Information Technologies* 27 (2022).
- [10]. F. S. Kia, S. D. Teasley, M. Hatala, S. A. Karabenick, and M. Kay, How patterns of students dashboard use are related to their achievement and self-regulatory engagement, *LAK 2020 Conference Proceedings – Celebrating 10 years of LAK: Shaping the Future of the Field – 10th International Conference on Learning Analytics and Knowledge*, 340 (2020).
- [11]. I. Jivet, M. Scheffel, H. Drachsler, and M. Specht, Awareness is not enough. pitfalls of learning analytics dashboards in the educational practice, *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning* 10474, 12, (2017).
- [12]. B. Kitchenham and P. Brereton, A systematic review of systematic review process research in software engineering, *Information and Software Technology*, 2049 (2013).
- [13]. H. Zhang, M. A. Babar, and P. Tell, Identifying relevant studies in software engineering, *Information and Software Technology* 53 (2011).
- [14]. P. O. Silva-Vasquez, V. Y. Rosales-Morales, and E. Benitez-Guerrero, Automatic code generation of user-centered serious games: A decade in review, *Program. Comput. Softw.* 48, 685–701 (2022).
- [15]. S. I. Fernandez Gregorio, L. G. Montane-Jimenez, C. M. Godoy, and V. Y. Rosales-Morales, Architecture for groupware oriented to collaborative medical activities in the rehabilitation of strokes, *Program. Comput. Softw.* 49, 643–656 (2024).
- [16]. J. A. Reyes, The skinny on big data in education: Learning analytics simplified, *TechTrends* 59, 75 (2015).

- [17]. R. Bodily and K. Verbert, Review of Research on Student-Facing Learning Analytics Dashboards and Educational Recommender Systems, *IEEE Transactions on Learning Technologies* 10, 405 (2017).
- [18]. P. Rojas-Castro, Learning Analytics: una revisi n de la literatura, *Educaci n y Educadores* 20, 106 (2017).
- [19]. D. T. Tempelaar, B. Rienties, and Q. Nguyen, Towards actionable learning analytics using dispositions, *IEEE Transactions on Learning Technologies* 10, 6 (2017).
- [20]. O. Viberg, M. Hatakka, O. B alter, and A. Mavroudi, The current landscape of learning analytics in higher education, *Computers in Human Behavior* 89, 98 (2018).
- [21]. C. Herodotou, B. Rienties, A. Borooowa, Z. Zdrahal, and M. Hlosta, A large-scale implementation of predictive learning analytics in higher education: the teachers' role and perspective, *Educational Technology Research and Development* 67, 1273 (2019).
- [22]. J. Guerra, M. Ortiz-Rojas, M. A. Zuniga-Prieto, E. Scheihing, A. Jimenez, T. Broos, T. De Laet, and K. Verbert, Adaptation and evaluation of a learning analytics dashboard to improve academic support at three Latin American universities, *British Journal of Educational Technology* 51, 973 (2020).
- [23]. S. Ranjeeth, T. P. Latchoumi, and P. V. Paul, A Survey on Predictive Models of Learning Analytics, *Procedia Computer Science* 167, 37 (2020).
- [24]. T. De Laet, M. Millecamp, M. Ortiz-Rojas, A. Jimenez, R. Maya, and K. Verbert, Adoption and impact of a learning analytics dashboard supporting the advisor – Student dialogue in a higher education institute in Latin America, *British Journal of Educational Technology* 51, 1002 (2020).
- [25]. C. Guzman-Valenzuela, C. Gomez-Gonzalez, A. Rojas-Murphy Tagle, and A. Lorca-Vyhmeister, Learning analytics in higher education: a preponderance of analytics but very little learning? *International Journal of Educational Technology in Higher Education* 18, 23 (2021).
- [26]. A. Rafique, M. S. Khan, M. H. Jamal, M. Tasadduq, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, Integrating learning analytics and collaborative learning for improving student's academic performance, *IEEE Access* 9, 167812 (2021).
- [27]. C. J. Perez Sanchez, F. Calle-Alonso, and M. A. Vega-Rodr guez, Learning analytics to predict students' performance: A case study of a neurodidactics-based collaborative learning platform, *Education and Information Technologies* (2022).
- [28]. R. Kaliisa, B. Rienties, A. I. M rch, and A. Kluge, Social learning analytics in computer-supported collaborative learning environments: A systematic review of empirical studies, *Computers and Education Open* 3 (2022).
- [29]. H. Huang, L. Jew, and D. Qi, Take a MOOC and then drop: A systematic review of mooc engagement pattern and dropout factor, *Heliyon* 9, e15220 (2023).
- [30]. P. Prinsloo, M. Khalil, and S. Slade, Learning analytics as data ecology: a tentative proposal, *Journal of Computing in Higher Education*, 1 (2023).
- [31]. K. Kaur and O. Dahiya, Role of educational data mining and learning analytics techniques used for predictive modeling, in *2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM)* (2023) pp. 1–6.
- [32]. M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Th s, A reference model for learning analytics, *Int. J. Technol. Enhanc. Learn.* 4, 318–331 (2012).
- [33]. F. Gutierrez, K. Seipp, X. Ochoa, K. Chiluita, T. De Laet, and K. Verbert, Lada: A learning analytics dashboard for academic advising, *Computers in Human Behavior* 107, 105826 (2020).
- [34]. J. Kim, I. Jo, and Y. Park, Effects of learning analytics dashboard: Analyzing the relations among dashboard utilization, satisfaction, and learning achievement, *Asia Pacific Education Review* 17, 13 (2016).

Информация об авторах / Information about authors

Ангел САЛАС-МАРТИНЕС получил степень магистра в области информационных сетей и интегрированных систем в Национальной лаборатории передовых вычислений (LANIA) штата Веракрус (Мексика). В настоящее время обучается в аспирантуре Университета штата Веракрус. Область научных интересов: программная инженерия, учебная аналитика (LA), панель мониторинга аналитики обучения (LAD), интеллектуальный анализ данных (DM), интеллектуальный анализ данных в образовании (EDM).

Angel SALAS-MARTINEZ – holds a Master's degree in Networks and Integrated Systems from National Laboratory of Advanced Computing (LANIA) in Veracruz, Mexico. He is a PhD student

in Computer Science at the Universidad Veracruzana in Mexico. His areas of interest are Software Engineering, Learning Analytics (LA), Learning Analytics Dashboard (LAD), Data Mining (DM) and Educational Data Mining (EDM).

Альберто РАМИРЕС-МАРТИНЕЛЬ имеет степень PhD по исследованиям в области образования, полученную университете Ланкастера (Великобритания); степень магистра по программированию и коммуникациям Университета прикладных наук (Фуртванген, Германия); степень по программной инженерии Национального университета Мексики, степень бакалавра по гуманитарным наукам Университета монастыря Сор-Хуана. В настоящее время работает профессором и исследователем в Университете штата Веракрус (Мексика). Его научные интересы в основном связаны с внедрением информационных и коммуникационных технологий в высших учебных заведениях.

Alberto RAMIREZ-MARTINELL – holds a PhD in Educational Research from Lancaster University, United Kingdom; a Master's degree in Computer Science and Media from the University of Applied Sciences in Furtwangen, Germany; a degree in Computer Engineering from Universidad Nacional Autónoma de México and a BA in Humanities from Universidad del Claustro de Sor Juana. He is currently Full-time Professor and Researcher at the Universidad Veracruzana, Mexico. His main research topic revolves around a disciplinary approach for the incorporation of Information and Communication Technology (ICT) in Higher Education Institutions.

Самюэль МАРТИНЕС-РАМОС – профессор Высшего технологического института в Пероте Национального технологического центра Мексики (TecNM/ITSP). Получил степень магистра в области информационных сетей и интегрированных систем в Национальной лаборатории передовых вычислений (LANIA) Веракруса (Мексика). Сфера научных интересов: разработка технологий, бизнес-аналитика, Интернет вещей.

Samuel MARTINEZ-RAMOS – holds a Master's degree in Networks and Integrated Systems from National Laboratory of Advanced Computing (LANIA) in Veracruz, Mexico. He is currently a Full Time Professor at the Tecnológico Nacional de México / Instituto Tecnológico Superior de Perote. His areas of interest are technology development, Business Intelligence and Internet of Things.