Informatics in Education, 2020, Vol. 19, No. 3, 361–397
© 2020 *Vilnius University, ETH Zürich*
DOI: 10.15388/infedu.2020.17

Monitoring Academic Performance Based on Learning Analytics and Ontology: A Systematic Review

Laecio Araujo COSTA[1], Leandro Manuel PEREIRA SANCHES[2], Ricardo José ROCHA AMORIM[3], Laís do NASCIMENTO SALVADOR[2], Marlo Vieira dos SANTOS SOUZA[2]

[1]*Federal Institute of Sertão Pernambucano, Informatics Collegiate – Petrolina/Pe, Brazil*
[2]*Federal University of Bahia, Computer Science Graduate Program – Salvador/Ba, Brazil*
[3]*State University of Bahia, Department of Education, Senhor do Bonfim/Ba, Brazil*
e-mail: laecio.costa@ifsertao-pe.edu.br, {leandrompsanches, amorim.ricardo}@gmail.com, {laisns, msouza1}@ufba.br

Received: December 2019

Abstract. This paper presents a systematic literature review of the coordinated use of Learning Analytics and Computational Ontologies to support educators in the process of academic performance evaluation of students. The aim is to provide a general overview for researchers about the current state of this relationship between Learning Analytics and Ontologies, and how they have been applied in a coordinated way. We selected 31 of a total of 1230 studies related to the research questions. The retrieved studies were analyzed from two perspectives: first, we analyzed the approaches where researchers used Learning Analytics and Ontologies in a coordinated way to describe some Taxonomy of Educational Objectives; In the second perspective, we seek to identify which models or methods have been used as an analytical tool for educational data. The results of this review suggest that: 1) few studies consider that student interactions in the Learning Management System can represent students' learning experiences; 2) most studies use ontologies in the context of learning object assessment to enable learning sequencing; 3) we did not identify methods of evaluation of academic performance guided by Taxonomies of Educational Objectives; and 4) no studies were identified that report the coordinated use of Learning Analytics and Computational Ontologies, in the context of academic performance monitoring. Thus, we identify future directions of research such as the proposal of a new model of evaluation of academic performance.

Keywords: distance education and online learning, educational objectives, learning analytics, ontology.

1. Introduction

The evolution of Information and Communication Technologies (ICT) has made it possible to tackle challenges regarding storing and processing huge amounts of data in the educational context so as to promote gains in the learning process. The growing use of learning platforms in Distance Education is an example of this. Unlike classroom teaching, e-Learning courses have specific characteristics such as the transactional distance between actors (educators and students) and the use of learning platforms, called as Learning Management System (LMS) or Massive Open Online Courses (MOOC). This transactional distance presents challenges in the teaching process such as: i) the lack of information about the students' real academic progress, ii) the attempt to predict the result of the students' academic performance, iii) the difficulty in making pedagogical decisions due to the low support of Information Systems, iv) the difficulty in keeping the student engaged, and v) high dropout rates (Yago *et al.*, 2018; Villagra-Arnedo *et al.*, 2017; Iglesias-Prada *et al.*, 2015). Thus, researchers are making an effort to minimize such challenges using computational resources applied to the educational context.

In e-Learning courses (also referred to as web-based education), learning takes place through students' interactions with the pedagogical support resources (Chat, Forum, Wiki, Pages, Links, and others) available in the learning platforms. Muñoz *et al.* (2015) describe LMS as a work environment used to support content management, the academic process, and the monitoring of learning development from the data generated during student interactions for knowledge building.

A MOOC platform usually refers to courses which are "massive, with theoretically no limit to enrolment; open, allowing anyone to participate, usually at no cost; online, with learning activities typically taking place over the web; and a course, structured around a set of educational goals in a defined area of study" (Educause, 2013, p.1). Being "massive" and "open", these courses are designed to be accessible to many more learners than would be possible through conventional teaching. They are often free of charge and participation need not be limited by the geographical location of the learners.

However, for these learning platforms, the scale, heterogeneity, and distributed nature of the students requires new methods for both providing student support (engagement) and guiding teacher intervention based on students' academic performance. These characteristics undermine the effectiveness of traditional methods such as direct observation or the use of questionnaires, or interviews (Alraimi *et al.*, 2015; Margaryan *et al.*, 2015). This teaching modality focuses on providing interactive learning environments, encouraging discussions, social network engagement, peer assessment, and teaching based on educational objectives.

Thus, educators look for computational tools allied to the learning platforms which enable the use of educational data analysis techniques to assist in the assessment of students' academic performance and to support in pedagogical decisions, promoting a more personalized learning experience and the construction of student knowledge through their learning experiences.

In order to build students' knowledge, educators plan educational objectives and use the technological resources available through the learning platforms (Webcast, Forum, Evaluative Activities, Chat, Wiki, Document Page, Learning Object Repository, among others) to promote learning gains. These resources assist the teacher in the process of planning, creating, controlling and managing the course or discipline online.

Bloom *et al.* (1956) state that the planning of educational goals is an intrinsic activity to the teaching process. Each and every academic activity has at least one planned educational objective. He presented a Classification of Educational Objectives, called Bloom's Taxonomy. It has a hierarchical structure that aims to assist the educator in planning the objectives of the class. Bloom's Taxonomy is described by verbs (actions intended by the educator) distributed in six hierarchical levels that relate to the student's level of knowledge. Bloom also reinforces that the educator can use the taxonomy to monitor and evaluate student learning. This evaluation has a diagnostic function, which makes it possible to verify the situation of students' learning to propose new means of mediation and intervention by the educator.

In the teaching process, assessment is used to read students' learning and the result may help to promote students' learning engagement and self-regulation (Pelissoni, 2009). Lukesi (2011) reinforces that the learning assessment process is a means of making the acts of teaching and learning productive and satisfying, contributing to the analysis and decision of which pedagogical actions should be taken.

Assessing learning enables the educator to diagnose students regarding the acquisition of planned skills and competences. According to Goodyear and Retalis (2010), in the e-Learning mode, this acquisition occurs through interactions (student-content, student-teacher, and student-environment) with the resources available on the learning platform and that produce a large amount of educational data. Assuming that students produce their own educational data and that it is stored in the learning platform database, we seek to investigate which available computing resources can enable the processing and analysis of these data.

In order to explore educational data and, consequently, to improve student success, a method was defined called Learning Analytics (LA). Learning Analytics is derived from methods of educational data mining to reveal patterns applied to the learning flow. Siemens and Long (2011, p.34) define LA as "the use of intelligent data (student-produced data) and analytical models to uncover information and social connections as well as predict and advise on learning". The Society for Learning Analytics Research (SoLAR) define LA as "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs" (SoLAR, 2011). Learning analytics is large implements to improving learner success.

There are many learning platforms, MOOCs and LMS, in which LA can be applied. These learning platforms provide massive amounts of data about the learning flow of the learners and the way in which they interact with the online learning environment. For Taraghi *et al.* (2014) these massive amounts of educational data about the students learning experiences available in the learning platform, indicate a high potential to use LA.

The integration of LA with a classification structure of educational objectives allows the consistent assessment of students' academic performance. Thus, the perspective to analyze educational data on the viewpoint on a Taxonomy of Educational Objectives, it is necessary that software agents understand this taxonomy structure. In this case, Computational Ontologies can describe the taxonomy through the formal representation of abstract concepts and properties, besides being able to infer knowledge about the represented information. Gruber (1993) defines an ontology as a formal explicit specification of a shared conceptualization. There are several studies in which the field of education has been represented through ontologies with encouraging results (Bourdeau *et al.*, 2007; Al-Yahya *et al.*, 2015; Psyche *et al.*, 2005; Vesin *et al.*, 2012; Korchi and Abdellah, 2015; Quinn *et al.*, 2017; Amorim *et al.*, 2006).

In this context, we seek to evaluate and interpret the studies available in the literature related to the following main research question: "How can Learning Analytics and Computational Ontologies help to monitor learning based on a Taxonomy of Educational Objectives? ". A Systematic Literature Review (SLR) was carried out to gather primary studies to assist in the search by evidences and the development of future research. We define two objectives for this work: i) to identify the main methods for monitoring academic performance based on Learning Analytics and Learning Objectives Taxonomies formalized by ontologies, and ii) to identify gaps and future opportunities to conduct research and create tools to advance the field of academic performance monitoring. This systematic review followed the manual by Kitchenham and Charters (2007), and Peterson *et al.* (2003).

This paper is structured as follows: the background is described in Section 2, Section 3 presents the Systematic Literature Review process and methodology, the results and analysis are presented in Section 4, the main findings and their respective analyzes are discussed in Section 5. Section 6 presents threats to the validity of this research, and finally Section 7 presents a summary of the work and directions for future work.

2. Background

The educational system has been increasingly using applied research to improve it through the application of available technological resources. Nowadays, learning platforms, for instance, LMS and MOOCs, often support the modalities of the classroom and non-classroom teaching. These platforms, assist educators and students in the management of the teaching/learning process (Castro *et al.*, 2007). It produces large data sets about students and your learning experiences. Extracting useful information from this mass of data has attracted the interest of researchers in the field of educational data analysis and online education. Some researchers positively correlate this wide range of educational data with the student engagement (Campbell *et al.*, 2007), academic achievement (Macfadyen and Dawson, 2012), and learning outcomes (Archer *et al.*, 2014; Hrastinski, 2009).

The growing demand for analysis of these massive amounts of educational data has strengthened the convergence of some lines of research (Big Data, Data Mining, and Analytics) in the educational context by assisting in the application of LA methods. For Elias (2011, p. 5), "Learning Analytics seeks to increase analytical skills, predict behaviour, predictive pedagogical action, and then feedback the education system with these outcomes to improve predictions over time of the learning".

Khalil and Ebner (2015) proposed a method that describes an LA life cycle. The authors discuss the essence, objectives, and methodologies of LA and propose a first prototype that describes its entire process. They present a reference model (Fig. 1) for the application of the LA in a cyclical process that involves four main parts: Learning environment where stakeholders produce data; Big Data, which consists of massive numbers of datasets; Analytics, which comprises different analytical techniques; Act, where objectives are achieved to optimize the learning environment. According to the same authors, LA is a promising research field, which provides tools and platforms that support researchers in Technology Enhanced Learning.

However, the tools of supervising students available in learning platforms are not easy to use and do not enable the consistent assessment of student learning progress. According to Yago *et al.* (2018), educators try to use the tools available on the learning

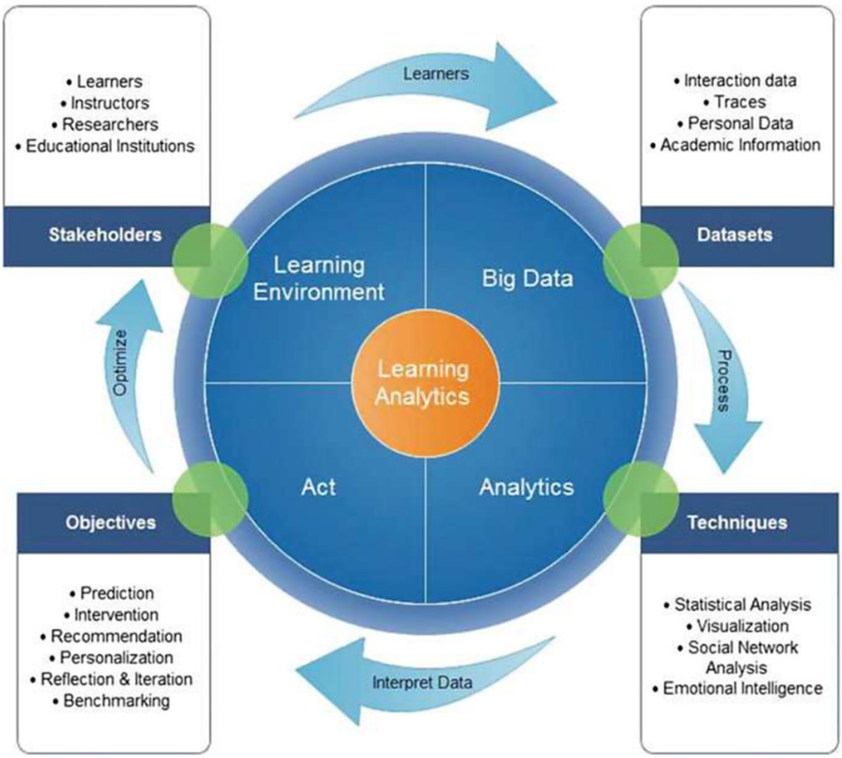

Fig. 1. Reference Method for Learning Analytics Application.
Khalil and Ebner (2015).

platforms to supervise and evaluate students' learning progress in distance education, however, they do not obtain the necessary information to evaluate the students' academic performance.

Vilagra-aredo *et al.* (2018) claim that modern educational theories advocate for a student-centered teaching process with a truly formative assessment. Therefore, for an academic performance assessment model to be useful to the educator, it must provide substantial data about student learning, provide tools for the educator to properly interpret progress and detect trends and patterns.

Due to the transactional distance in this type of learning, Lima and Fialho (2011) state that the didactic and pedagogical organization of the courses offered must be planned so that each student can organize and build an autonomous learning process. To carry out the pedagogical planning of a subject, educators use and define the educational objectives that they intend to achieve with that planning.

According to Bloom *et al.* (1956), Anderson and Krathwohl (2001), Lima and Fialho (2001), and Freire (2000), Educational Objectives are pedagogical resources inherent to the teaching process, which guide the planning of the course and allow the assessment of student academic performance, as well as the monitoring of compliance with planned goals. Haydt (2011) defines educational goals as desired and anticipated outcomes for an educational action. Educational results are sought by the educator through the pedagogical activities. Any decision or pedagogical action is supported by educational goals and having clear goals is the first and perhaps the "most important principle" for designing an ideal ODL-based course (Pelissoni, 2009).

Fernandez-Delgado *et al.* (2014) report that student learning activities are related to one or more educational goals and that an educational goal is achieved through one or several learning activities. If a particular learning activity is successfully completed by a student, the educational objectives attached to that activity must be achieved. Thus, information about the level of academic performance that students present with the fulfilment of educational objectives, through the application of learning activities, allow teachers to introduce new pedagogical actions in order to promote student performance.

In 1956 Benjamin Bloom defined an Educational Objectives classification structure divided into six hierarchical levels. Each level has a set of verbs that describe objectives (actions) to be accomplished by students at each level of the taxonomy. Bloom's Taxonomy was pioneering, it is widely referenced and used to assist educators in planning educational activities. The use of a taxonomy in the processing of educational data may enable a consistent assessment of student progress. Illustrative examples of taxonomies are Bloom's (Bloom *et al.*, 1956), Revised Bloom's Taxonomy (Krathwohl and Anderson, 2001), and SOLO (Biggs and Collis, 1982). However, Bloom's taxonomy is the most referenced and widely used.

For software agents to understand the taxonomic structures of the available Educational Objectives, it is necessary to formalize taxonomies through Ontology Engineering. Ontology has been applied in the educational context for several purposes, for example to relate Learning Objects with the sequencing of student learning (Lima *et al.*, 2017), to make semantic annotation of Learning Objects (Sanches *et al.*, 2017), to extract speech information (Zhang and Zhang, 2010) among others.

Recently, some specifications for maintaining LA interoperability have been transformed into ontologies, such as xAPI[1] or IMS Caliper[2]. These specifications allow to capture data from interactions in the learning platform and store it in a repository.

According to Yago *et al.* (2018), despite the common use of LA tools, there are no flexible monitoring or diagnostic approaches that can be applied to platform-independent LMS-supported courses providing feedback as a strategy for instruction / learning. From this perspective, educators seek to use different instruments and tools to partially supervise (due to the limitation of tools) the progress of students' academic performance. This work is an extension and update of the mapping performed by Costa *et al.* (2018).

3. Systematic Review Process

Systematic Literature Review is a form of secondary study that uses a well-defined methodology to identify, analyze and interpret all available evidence related to specific research questions that is impartial and (to some extent) repeaTable (Kitchenham and Charters, 2007). In this paper, we searched for publications that present the application methods of LA, Computational Ontologies, Educational Objectives Taxonomies and their relationship with academic performance monitoring process. In Section 3.1, we describe the methodology used in this study including the main research question and other secondary questions, inclusion and exclusion criteria and the data extraction process. Section 3.2 presents the data extraction process and quality assessment of the retrieved studies.

3.1. *Research Methodology*

Systematic Literature Review is a means used to identify, evaluate and interpret relevant literature related to research issues, topic area or phenomenon (Kitchenham and Charters, 2007). The main objective of this review is to gather primary studies that can help to draw conclusions about methods for monitoring and evaluating academic performance based on Taxonomies of Educational Objectives.

To perform this SLR we used the protocol and guidelines proposed by Kitchenham *et al.* (2009) and Dermeval *et al.* (2017). The SLR execution process can be grouped into three main phases, namely: SLR Planning, Execution, and Reporting. These steps consist of: i) formulating research questions; i) performing a comprehensive and exhaustive search for primary studies; iii) evaluating the quality of the included studies; iv) identifying and extracting the data necessary to answer the research questions; v) summarizing and synthesizing the study results; vi) interpreting the results to determine their applicability; and finally, vii) writing reports.

---

[1] xAPI - is an e-Learning specification that allows you to collect data about a wide range of experiences a student has, either through online or offline training (`http://www.xapi.com`).

[2] Caliper - enables institutions to collect learning resources from digital resources to better understand educational data and present information to students, instructors, and counsellors in meaningful ways (`https://www.imsglobal.org/activity/caliper`).

3.1.1. *Research Questions and Inclusion and Exclusion Criteria*

Carrying out an SLR requires one or more well-formulated and clear research questions (Sampaio and Mancini, 2007). The research questions were formulated according to the PICOC method (Population, Intervention, Comparison, Outcomes and Context), defined by Kitchenham and Charters (2007). Table 1 presents the adequacy of the PICOC method to define the research questions in this paper.

Gathering information from the PICOC method, we define the following main research question:

How can Learning Analytics and Computational Ontologies assist educators in monitoring learning based on a Taxonomy of Educational Objectives?

Based on the primary research question, we formulated a set of five secondary questions (RQ), listed in Table 2, to understand the methods, techniques, ontologies, tax-

Table 1
Adequacy of the PICOC model for the definition of the research question

Aspect	Meaning	Scope
Population	What? Or who?	E-learning students
Intervention	How?	Coordinated use of Learning Analytics and Computational Ontologies
Comparison	Compare with what?	Not applicable
Outcome	What you want to do	Academic performance monitoring guided by a Taxonomy of Educational Objectives
Context	In which context?	Educational

Source: Adapted from Ghani and Yasin (2013)

Table 2
Secondary Research Questions and their Descriptions

Research Question	Description
RQ1. Which methods allow you to evaluate academic performance based on Learning Analytics and Computational Ontologies in a coordinated way?	This question aims to provide an understanding of the approaches about LA and Ontologies to support in the assessment of the academic progress.
RQ2. Which ontologies assist in the academic performance monitoring process?	We intend to identify which computational ontologies are used in the context of this research.
RQ3. What elements of LMS are used as learning indicators?	This question identifies the learning indicators available in the EMS, which make it possible to extract information about students' experiences.
RQ4. What computational resources, techniques, or methods are used in the learning performance assessment process?	This question allows us to identify techniques, methods or algorithms that have been developed / applied within the scope of this research.
RQ5. What are learning objective classification hierarchical structures used to monitoring student academic progress?	This question will highlight the taxonomies of Educational Objectives used.

Table 3
Inclusion and Exclusion Criteria Applied in the Systematic Literature Review

Inclusion criteria (IC)	Exclusion Criteria (EC)
IC1. Primary studies IC2. Studies that presents methods and practical aspects of using Techniques Learning Analytics and Learning Taxonomies in LMS IC3. Articles covering the development and evaluation of experimental studies involving Learning Analytics and Computational Ontologies IC4. Papers about Educational Objectives and LMS IC5. Papers about conceptual models in the context of this research	EC1. Article in a language other than English EC2. Article not available in digital library EC3. Technical reports, documents in the form of summaries, as well as secondary literature reviews EC4. Studies that address only philosophical aspects EC5. In case of duplicate studies, consider more current EC6. Redundant Studies by the same author.

onomies and their applications in the context of this research and to identify possible research gaps in this area.

To select the retrieved studies, we applied the following inclusion and exclusion criteria presented in Table 3.

3.1.2. *Classification Scheme*

The classification of the studies was based on four directions: (i) meta-information of the studies, ii) the analysis of studies relating LA and Ontologies, (iii) identification of academic performance indicators in the LMS, and (iv) the point of view of monitoring academic performance, studies guided by a Taxonomy of Educational Objectives were considered. To answer the research questions, we use the classifications described below:

- **Meta-Information**: We searched for general characteristics of the retrieved studies – type of publication (Journal, Conference or Workshop), research method (case study, survey, controlled experiment, etc.) and the temporal distribution of publications.
- **Learning Analytics and Ontologies** (RQs 1, 2, and 4): We investigated methods and techniques that enable the coordinated application of LA and Ontologies to monitor students' academic performance. We made a more detailed analysis in search of the main resources used. In terms of ontology, we investigated which ontological representations were developed or used in the field of Taxonomies of Educational Objectives and their relationship with LA.
- **Performance Indicators** (RQ 3): We analyzed the selected studies to investigate which mechanisms of interaction in the LMS provide indicators that allow the monitoring of academic performance.
- **Taxonomy of Educational Objectives** (RQ 5): We sought to investigate which taxonomies are most referenced and used in the context of this research.

3.2. *Conducting the Literature Systematic Review*

This section presents how the primary studies were selected and extracted while conducting this Systematic Literature Review.

Digital libraries (research sources) were used to consult the primary works to be studied. Digital libraries allow (semi)automatic searching using the available search engine. The choice of digital libraries was defined according to Costa *et al.* (2015) and from the qualitative analysis performed by Buchinger *et al.* (2014), which used an initial sample of 40 academic search sources. The five most cited libraries were used in this research, namely:

- IEEE Xplore
- ACM Digital Library
- Science Direct
- Springer Link
- Scopus – Elsevier

3.2.1. *String Search Formulation*

The search string was defined from keywords related to the research questions. The keywords were defined according to Table 4.

We performed keyword tests to observe the different results in the search engines and identify possible spelling variations. Table 5 describes the search strings used in the assay and their results when applied to the search engines from the sources cited in this Section.

As presented in Table 5, we tested five different search strings to retrieve a more representative amount of work related to the object of this research. String 1 has been formulated for an overview of the amount of work related to the terms of education. String 2 is more specific and aims to retrieve relevant studies that relate LA and Ontologies to educational terms, but the result was not satisfactory.

Table 4
Terms extracted from the search question for search string definition

Context	Keyword
Computational	*"Learning Analytics"* *"Educational Data Mining"* *Ontology*
Educational	*"Educational Objective"* *"Instructional Objective"* *"Learning Objective"* *"Cognitive process"* *"Cognitive learning"* *"Cognitive objective"* *"Educational theory"* *"Learning theory"*

Table 5

Search String Calibration Testing

String	Description	Total
String 1	"educational objective" OR "instructional objective" OR "learning objective" OR "cognitive objective" OR "cognitive process" OR "cognitive learning" OR "educational theory" OR "learning theory"	207.059
String 2	"learning analytics" AND ontology AND ("educational objective" OR "instructional objective" OR "learning objective" OR "cognitive objective" OR "cognitive process" OR "cognitive learning" OR "educational theory" OR "learning theory")	0
String 3	ontology AND ("educational objective" OR "instructional objective" OR "learning objective" OR "cognitive objective" OR "cognitive process" OR "cognitive learning" OR "educational theory" OR "learning theory")	749
String 4	("learning analytics" OR "educational data mining") AND ontology AND ("educational objective" OR "instructional objective" OR "learning objective" OR "cognitive objective" OR "cognitive process" OR "cognitive learning" OR "educational theory" OR "learning theory")	0
String 5	("learning analytics" OR "educational data mining") AND ("educational objective" OR "instructional objective" OR "learning objective" OR "cognitive objective" OR "cognitive process" OR "cognitive learning" OR "educational theory" OR "learning theory")	481

String 3 aims to retrieve studies that relate the term Computational Ontologies to other terms in the educational field. This string aims to answer specifically the research question RQ2 (Which ontologies assist in the academic performance monitoring process?). String 4 is an extension of string 2, both returned no results. String 5 is intended to retrieve studies that relate the terms Learning Analytics or Educational Data Mining to terms in the education field. This string obtained a satisfactory result in relation to sets 1, 2, and 4.

The performed tests identified the keywords which provided the best representation of the studies related to the object of this research. As string 2 and 4 did not obtain results, we used the two other strings (3 and 5).

3.2.2. *Search Process*

This Systematic Review was carried out between March and July 2019. The process of collecting the articles using strings 3 and 5, applying the inclusion and exclusion criteria, and the selection of the study was conducted in four stages, as presented in the following sections. Table 6 presents the set of retrieved studies classified according to the Digital Library from where they came using the selected strings.

The process of extraction and selection of relevant works for follow-up of this research was performed according to the steps presented in Fig. 2.

In the initial stage, we retrieved 1230 publications by applying both strings to digital library search engines, distributed as follows: ACM (151), IEEE Xplore (34), Science

Table 6
List of retrieved articles by applying search strings to selected Digital Libraries

BASE	Retrieved Articles	
	String 3	String 5
ACM	106	45
IEEE	30	4
SCIENCE DIRECT	26	162
SCOPUS	214	86
SPRINGER	373	184
Total	**749**	**481**
Grand Total	**1230**	

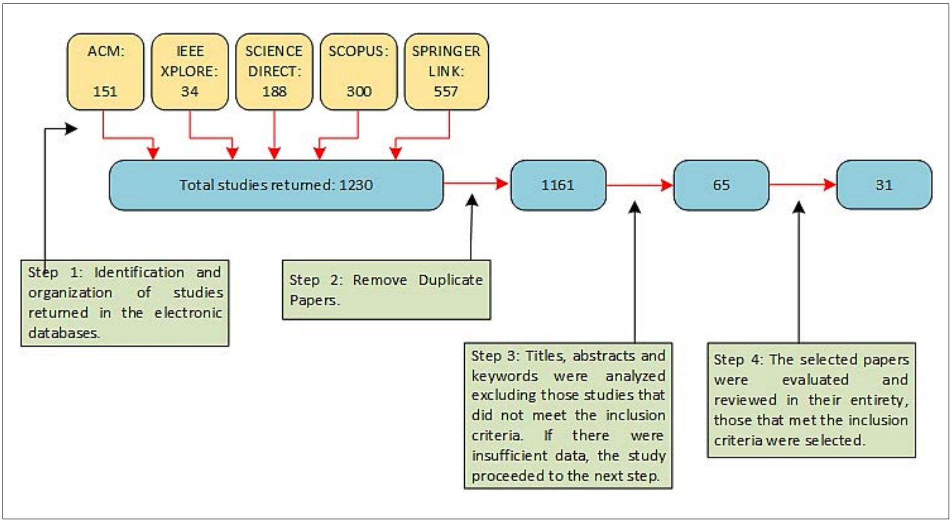

Fig. 2. Collection and processing flowchart of studies retrieved using strings 3 and 5 in digital libraries. Adapted from Dermeval et al. (2017).

Direct (188), Scopus (300), and Springer Link (557). All articles were managed and organized with the support of the StArt tool (LaPES, 2013).

In the second stage, 69 duplicate publications were identified and among them the most current publications were considered. Then the inclusion and exclusion criteria were applied by analyzing the title, keywords and summary of the 1161 articles, resulting in 65 studies (third stage).

In the fourth stage, the 64 works were completely read in order to evaluate and extract the data necessary to answer the research questions, excluding papers that are not related to the object of this research. At the end of this stage, 31 works resulted (2.43% of the total). Tables 7 and 8 represent the process of selecting the jobs retrieved by applying strings 3 and 5, respectively.

Table 7
List of articles included after applying the inclusion and exclusion criteria for string 3

Digital Library	Step 1	Step 2	Step 3	Step 4	Included
ACM	106	3	103	3	2
IEEE	30	-	27	3	1
SCIENCE DIRECT	26	-	24	2	0
SCOPUS	214	30	177	3	1
SPRINGER	373	9	362	2	2
Total	**749**	**42**	**693**	**13**	**6**

Table 8
List of excluded articles after applying the inclusion and exclusion criteria for string 5

Digital Library	Step 1	Step 2	Step 3	Step 4	Included
ACM	45	1	33	11	3
IEEE	4	-	3	1	1
SCIENCE DIRECT	162	1	145	6	10
SCOPUS	86	36	19	33	6
SPRINGER	184	1	173	10	5
Total	**481**	**39**	**373**	**44**	**25**

3.2.3. *Quality Assessment*

Assessing the quality of primary studies is necessary to limit bias in performing the systematic review and guide interpretation of findings (Kitchenham and Charters, 2007; Higgins and Thomas, 2008). In addition to the general inclusion / exclusion criteria, it is considered essential to evaluate the “quality” of primary studies. This assessment is intended to:

- Provide even more detailed inclusion / exclusion criteria.
- Investigate whether quality differences provide an explanation for differences in study results.
- Consider the importance of individual studies when results are being synthesized.
- Guide the interpretation of the findings and determine the strength of the inferences.
- Guide recommendations for further research.

Evaluations are usually based on “quality instruments”, that is checklists of factors that need to be evaluated in each study (Kitchenham and Charters, 2007). The quality evaluation of the selected studies was obtained using scoring technique that evaluates the credibility, integrity and relevance of each study based in Dermeval *et al.* (2017).

In general, the “quality” of a study is closely linked to the research methods used and the validity of the findings generated by the study. Quality refers to the conduct and analysis of primary studies that are likely to avoid systematic errors or bias. Biased

Table 9
Study quality checklist

Nº	Question	Definition
CQ1	Is there justification for conducting the study? (Mahdavi-Hezavehi *et al.*, 2013)	Description
CQ2	Is there a clear statement of the research objectives? (Dyb and Dingsyr, 2008)	Description
CQ3	Is the proposed technique clearly described? (Achimugu *et al.*, 2014)	Description
CQ4	Was the study empirically evaluated? (Ding *et al.*, 2014)	Verification
CQ5	Has an adequate description of the applied test occurred? (Dyb and Dingsyr, 2008), (Mahdavi-Hezavehi *et al.*, 2013)	Verification
CQ6	Have all methods and tools been fully defined? (Kitchenham and Charters, 2007)	Verification
CQ7	Can the proposed method be applied in other contexts? (Ding *et al.*, 2014)	Applicability
CQ8	Were the results satisfactory? (Ding *et al.*, 2014)	Applicability
CQ9	Is the document research-based (or is it merely a “lessons learned” report based on expert opinion)? (Kitchenham and Charters, 2007)	Applicability
CQ10	Do the researchers explain the limitations or any problems with the validity / reliability of the method used? (Kitchenham and Charters, 2007)	Applicability

primary studies are more likely to provide misleading results, they are also likely to generate misleading systematic analyzes.

Thus, the quality analysis was based and adapted from Kitchenham and Charters (2007), Dyba and Dingsyr (2008), Mahdavi-Hezavehi *et al.* (2013), and Achimugu *et al.* (2014). The questions of verification of the quality of the studies, presented in Table 9, use three classifications:

- **Description:** Refers to the way the method is described in the text of the article and the level of detail.
- **Verification:** It focuses on the description of the tests applied to validate the proposed method.
- **Applicability:** This analyzes if the method has been validated and can be used in different contexts.

Each of the ten criteria (Table 9) was scored on a ternary scale: “1” awarded to a study when a question may be answered “Yes”, “0” when the answer is “No” and “0.5” if the answer is “Partially”. If the answer to criteria 1, 2 or 3 is “No” then the quality of the study under review should not be continued.

3.2.4. *Data Extraction*

In order to guide this process of data extraction, we adopted the guidelines of Kitchenham and Charters (2007). At this stage of the process we use a data extraction form (Table 10) designed to gather the expected information as per characteristics described in Section 3.1.1 (*Research questions and Inclusion and exclusion criteria*). Using the data extraction form enables us to record information from studies under review and to specify how each study is related to our research questions.

Table 10
Extraction Form

#	Study Data	Description	Relevance
1	Study Identifier	Unique ID for the study	Study Overview
2	Extraction Date	-	Study Overview
3	Authors, Year and Title	-	Study Overview
4	Article Source	Library in which to index the article	Study Overview
5	Post Type	Magazine, Conference, Workshop, Book Chapter, and etc.	Study Overview
6	Application Context	Industrial or Academic	Study Overview
7	Research method	Controlled Experiment, Case Study, Survey, etc.	Study Overview
8	Evidence	What evidence indicates the coordinated use of Learning Analytics and Computational Ontologies in learning assessment?	QP1
9	Ontological domain	Do you use any kind of ontological structure?	QP2
10	Indicators evaluated	What indicators in the EMS are used in the student performance appraisal process?	QP3
11	Contribution	Does it present any proposal / algorithm / technology to support learning assessment?	QP4
12	Classification of Educational Objectives	Do you refer to any classification of Educational Objectives?	QP5

Adapted from Dermeval et al. (2017)

4. Results and Analysis

A total of 31 studies met the inclusion criteria and their data were extracted. Before presenting the results and extracted data to answer each research question, we describe the results of the quality assessment and provide an overview of the characteristics of the studies.

Data were tabulated to present general information about the studies, such as identifier, authors, type of publication, context and research method. In addition to this information, we tabulate data on research questions and present graphs to provide a deeper view of multiple categories.

4.1. *Quality Assessment Result*

In order to increase the accuracy of the data extraction results, the quality of the selected studies was evaluated. The evaluation determines the validity of the inferences offered and verifies the credibility and synthesis of the results. Table 11 presents the results of the quality assessment according to the questions presented in Table 10 (Extraction form).

The data in Table 11 show that the quality does not vary greatly. There are studies of high quality as well as medium quality. The average quality assessment above 70%

Table 11
List of articles included in the review and their respective quality scores.
Studies are classified by identifier ID

ID		Authors	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total	%
STRING 3	P1	Ghannem (2014)	1	1	1	0	0	1	0	1	1	0	6.0	60
	P11	Ramesh *et al.* (2016)	1	1	1	1	1	1	1	1	1	0	9.0	90
	P125	Peters *et al.* (2017)	1	1	0	0	1	1	0	1	1	1	6.0	60
	P717	Guettat and Farhat (2016)	1	1	0	0	0	0	1	0	1	0	4.0	40
	P729	Fulantelli *et al.* (2014)	1	1	1	1	1	0	0	1	1	0	7.0	70
	P744	Antonelli *et al.* (2019)	1	1	1	1	0	1	1	1	1	0	7.5	75
STRING 5	S00	Gibson *et al.* (2014)	1	1	1	0	1	0	1	0	1	1	6.0	60
	S04	Khosravi *et al.* (2017)	1	0	1	1	1	0	1	1	1	1	7.0	70
	S09	Rienties *et al.* (2016)	1	0	1	1	1	0	1	1	1	0	6.5	65
	S46	Fernandez-Delgado *et al.* (2014)	1	1	1	1	1	1	1	1	1	0	8.0	80
	S61	Strang (2016)	1	1	1	1	1	1	1	1	1	1	10	100
	S63	Jayakodi *et al.* (2016)	1	1	1	1	1	1	1	1	1	0	8.5	85
	S66	Yamada *et al.* (2016)	1	1	1	1	0	1	1	1	1	0	7.5	75
	S99	Nussbaumer *et al.* (2012)	1	1	0	0	0	1	1	0	1	0	4.5	45
	S106	Pardo *et al.* (2016)	1	1	1	1	1	1	1	1	1	1	10	100
	S126	Wu and Wu (2018)	1	1	1	1	1	1	1	1	1	1	10	100
	S143	Tempelaar *et al.* (2014)	1	1	1	1	1	0	1	1	1	0	7.5	75
	S148	Hlioui *et al.* (2010)	1	1	1	0	0	1	1	1	1	0	6.5	65
	S183	Li *et al.* (2017)	1	1	1	1	1	1	1	1	1	0	9.0	90
	S208	Kostopoulos *et al.* (2017)	1	1	1	1	1	1	1	1	1	0	9.0	90
	S233	Dalipi *et al.* (2015)	1	1	1	1	1	1	1	1	1	0	7.5	75
	S319	Agudo-Peregrina *et al.* (2014)	1	1	1	1	1	1	1	1	1	1	10	100
	S326	Wanli *et al.* (2014)	1	1	1	1	1	1	1	1	1	1	9.0	90
	S336	Iglesias-Pradas *et al.* (2014)	1	1	1	1	1	1	1	0	1	1	9.0	90
	S349	Villagra-Arnedo *et al.* (2016)	1	1	1	1	1	1	1	1	1	1	9.5	95
	S354	Xing *et al.* (2016)	1	1	1	1	1	1	1	1	1	1	9.5	95
	S377	Kotsiants *et al.* (2010)	1	1	1	1	1	1	1	1	1	0	9.0	90
	S381	Yago *et al.* (2018)	1	1	1	0	0	1	0	1	1	0	5.0	50
	S421	Zacharis (2015)	1	1	1	1	1	1	1	1	1	0	9.0	90
	S473	Shorfuzzaman *et al.* (2019)	1	1	1	1	0	1	0	1	1	0	7.0	70
	S474	Aljohani *et al.* (2019)	1	1	1	1	1	1	0	1	1	0	8.0	80
Average			1.0	0.90	0.89	0.76	0.69	0.73	0.74	0.76	1.00	0.34	7.81	78.1
Standard deviation			0.0	0.27	0.31	0.43	0.44	0.40	0.41	0.36	0.00	0.47	1.68	16.77

indicates that the selected studies are coherent in the research elements and the data presented. The standard deviation measure is used to express a correlation coefficient between variables and here it indicates that the studies are close to the average.

The result of the quality assessment reflects the consistency among the studies identified in the review. If the quality assessment presents a result of studies with a high-quality rating, it is understood that future research will hardly change the observed effect. On the other hand, when the result has a very poor quality outcome, the estimates will probably change with the publication of new studies. Thus, the ten criteria used in the quality assessment represent an instrument that contributes to measure the credibility of the studies selected in the systematic review.

4.2. *Meta-Information*

4.2.1. *Overview of the Publication*

In this section, we present data from the collected and filtered primary studies. We map the general characteristics of the studies according to Section 3.1.2. The temporal distribution and type of publication of the retrieved studies are shown in Fig. 3.

Of the 31 selected primary studies, 18 (58.06%) were published at Conferences, 12 (38.71%) were published in Journals and 1 (3.23%) at Workshops. We observed that there was a significant growth in the number of publications from 2014 onwards. We should point out that there were no studies identified before 2010. As this review was conducted in the first half of 2019, the number of publications may still grow and surpass previous years, as shown in Fig. 3.

4.2.2. *Research Method*

The classification of research method was based on four categories (controlled experiment, case study, action research and comparative study) defined by Easterbrook *et al.* (2008). As shown in Fig. 4, the controlled experiment method constitutes the majority of the studies (18 studies, 58.06%), followed by the Case Study method (7 studies, 22.58%). The Action Research method was identified in 2 studies (6.45%), one study (3.22%) performed a Comparative Study and 3 studies (9.67%) did not mention the method used.

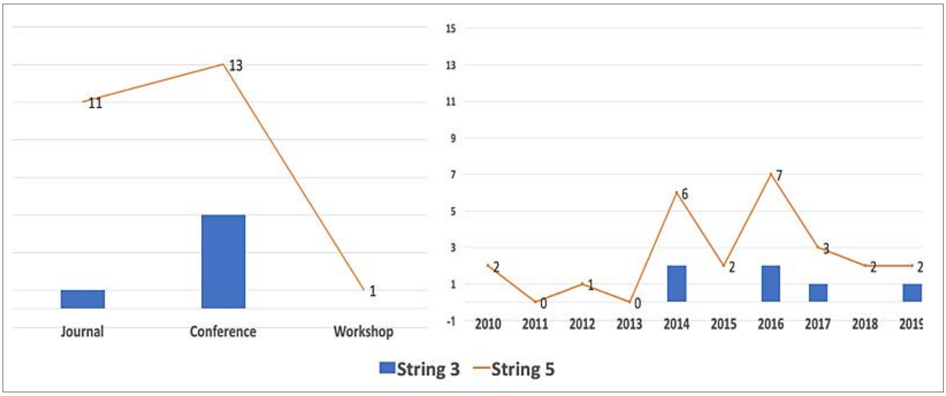

Fig. 3. Classification and temporal distribution of selected studies.

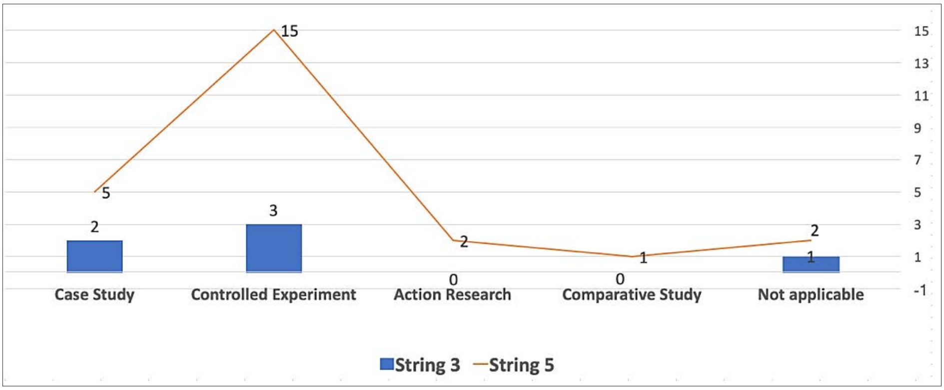

Fig. 4. Research method of selected studies.

The number of studies that conducted a controlled experiment indicates a significant increase in the development of tools that explore educational data using analytics to promote gains in the teaching-learning process.

4.3. *Learning Analytics and Ontology*

In this section, we present an analysis of the identified aspects of Learning Analytics and Ontology in order to answer the research questions RQ1, RQ2 and RQ4.

RQ1 – Which methods allow you to evaluate academic performance based on Learning Analytics and Computational Ontologies in a coordinated way?

RQ1's main objective is to identify and classify evidence on the use of LA and Computational Ontologies to assist in the academic performance assessment process by observing the achievement of educational objectives. Table 12 presents the classification of studies by resources used in the research.

The authors of studies P729, S126, S143, S381, S421, S473 and S474 generally cite that they use LA methods to assist in the analysis and processing of manipulated educational data, but in such studies, they do not highlight the use of LA and Ontology from the perspective of monitoring academic performance. The use of ontology was identified in studies P1, P11, P125, P717, P729, P744, S99 and S381. In studies P729 and S381 we identified the use of LA and Ontologies, but the purpose of these studies is not to track academic performance guided by a taxonomy of Educational Objectives.

Table 12
Evidence Mapping on the Use of Learning Analytics and Ontologies

Resource	Description	%
Learning Analytics	P729, S126, S143, S381, S421, S473, and S474	22.58%
Ontology	P1, P11, P125, P717, P729, P744, S63, S99, and S381	29.03%

In study P729, the authors apply LA methods in the context of mobile learning using MeLOD, a mobile learning environment that supports Linked Open Data (LOD). In the MeLOD environment, an ontological specification was used to represent all interactions that occur in the learning system. Study P729 focuses on the challenge of using LA techniques to support educational decision making in mobile learning environments.

The authors of study S381 present the ON-SMMILE model, a set of interconnected ontologies that combine information related to: i) students and their state of knowledge, ii) assessments that depend on rubrics and different types of objectives and iii) units of learning.

Among the selected works, no evidence of student academic performance monitoring was identified from the coordinated use of LA and Computational Ontologies techniques, guided by a Taxonomy of Educational Objectives.

RQ2 – Which ontologies assist in the academic performance monitoring process?

This research question aims to identify the main types of ontological structures that have been developed or used in research in the educational context. The classification of the ontological structures was performed after the extraction of the studies observing each type of structure explained by the authors. As presented in Section 3.2.1, a search string was specified to assist in the collection of studies that address this research question. Thus, with the execution of this SLR the use of Computational Ontologies was identified in 9 studies (P1, P11, P125, P717, P729, P744, S63, S99 and S381), Table 13.

In P1, the authors seek to assist teachers in the pursuit of Serious Games from the Learning Objectives related to the game. Based on the Learning Objectives and using a Game Ontology, the tool proposed by the authors classifies the Game according to the Bloom Taxonomy skill level. To define assessment metrics, the work is based on learning modeling theories, such as the IMS Learning Design specification, and uses Bloom's Taxonomy to define Game cognition levels.

The goal of P11 work is to automatically integrate Learning Objectives (LO) with Syllabus (course curriculum) from Domain Ontology to Data Structure courses. The authors

Table 13

Mapping of Studies on the Use of Computational Ontologies

Source	Study	Ontological Classification
String 3	P1	Domain Ontology – Game Ontology
	P11	Domain Ontology – Syllabus
	P125	Application Ontology – Medical Protocol-based Learning Tasks and Objectives
	P717	Domain Ontology for Student Profile
	P729	Top Ontology – DBPedia and Geonames
	P744	Application Ontology – CONALI
String 5	S63	Top Ontology – WordNet
	S99	Application Ontology – Self Regulated Learning
	S381	Application Ontology – ON-SMMILE

seek to capture relevant knowledge of Syllabus and OA by mapping it through an ontological representation. Thus, the competencies and cognitive level of the objective can be defined from Bloom's taxonomy. In P125, the authors proposed a tree-based Learning Objectives framework to shape and monitor the learning process. Article P717 aims to help students identify their Learning Objectives in Personal Learning (PLE), providing details to help unskilled pedagogy students formulate their personal learning goals.

Work P729 focuses on the challenge of using LA methods to support educational decision making in mobile learning environments. They use an approach that integrates Linked Open Data (LOD) in conjunction with pattern extraction techniques for interpreting insights resulting from student behavior. The goal is to assist teachers in tracking and evaluating students during mobile-based learning experiences. The framework is based on the relationships between the different types of interactions that occur in a mobile learning activity and the tasks, which are pedagogically relevant to the learning activities. The authors of this study use a mobile environment for LOD-based learning, MeLOD, to conduct the experiment.

In P744, a project to build an open networked platform for the learning of Industry 4.0 subjects was presented. The project will enable the creation of a laboratory using Virtual Reality (VR), where users can design and create an environment for training and simulation of industrial processes. Among the features available, this platform will enable students to customize their learning path through a modular approach. The platform is based on building blocks defined through the constructive alignment procedure. According to the authors, students will be able to co-create their learning path and learning content. The project uses VR resources, Industry 4.0 ontology and learning process ontology (CONALI) through Constructive Alignment. CONALI is an ontology that provides an understanding of learning units as the composition of a series of educational goals.

The authors of the S63 study use Natural Language Processing (PLN) techniques to extract verbs from evaluative questions, then evaluate the similarity of verbs using WordNet[3] (digital English verb base, classified as top ontology) and they classify verbs according to Bloom's taxonomy. In this article, the authors do not use a specific ontology to monitor academic progress, the objective is to classify the issues according to the taxonomy of Educational Objectives. The authors do not use LA methods, but claim that new information can be extracted from processing the resulting data through LA.

Study S99 presents a conceptual approach using learning ontology to assist in monitoring unobservable actions of students' cognitive and metacognitive activities. The approach deals with a modification of the self-regulated learning (SRL) cyclical model proposed by Zimmerman (2002). The SRL consists of four meta-cognitive and cognitive phases that occur during its process: learning process planning, resource search, real learning and reflection of the learning process. As the Personal Learning Environment (PLE) allows the creation of modules based on learning tools with a pedagogical approach, the assembly of pedagogical activities must follow content sequencing guidelines to enable more efficient learning. Thus, the authors suggest a model of mapping cognitive and meta-cognitive activities through a Learning Ontology that formalizes

[3] Lexical relational system for the English language, classified as Top Ontology. Available at: <https://wordnet.princeton.edu/>.

a Taxonomy of Learning Activities. The pedagogical perspective of the approach presented highlights the aspects of reflection and awareness of the learning process.

In S381, the authors present a theoretical model using a combination of ontologies to supervise student learning and recommend competency-based activities. The model is based on the semantic web to assist educators with decision making during the student learning process. The proposal combines education resources such as: IMS Learning Project, Student Model ontology, and learning classification.

RQ4 – What computational resources, techniques, or methods are used in the academic performance assessment process?

The purpose of this research question is to identify the learning platforms, techniques and methods used to monitor student performance. Regarding learning management platforms, Table 14 presents the learning environments that were identified among the retrieved studies.

Among the studies presented in Table 14, the following learning platforms stand out: MOOC (19.35%, 6 studies), Moodle (12.90%, 4 studies), other environments (SMEUS, MeLOD, and My Labs) represent (12.90%, 4 studies). More than half of the studies (54.83%, 17 studies) did not specify the type of platform used because the proposed architecture is generic so as to meet the requirements of the main learning environments.

Regarding the statistical methods and techniques used in the data analysis, we found that most studies focus on three techniques: Classification (Linear Regression), Clustering and Correlation Analysis, Table 15.

Table 14

Types of learning environments identified

Platform	Studies	Freq.	%
Not specified	S00, S09, S46, S63, S66, S106, S148, S208, S319, S349, S377, S381, S473, P1, P11, P125 and P729	17	54.83
MOOC	S04, S126, S183, S326, S354 and P744	6	19.35
Moodle	S61, S336, S421 and S474	4	12.90
Others	S143, S233, S473 and P717	4	12.90

Table 15

Statistical Techniques Identified

Technique	Studies	Freq.	%
Regression Analysis	S09, S61, S66, S106, S143, S183, S208, S319, S354 , S421	10	32.25
Data grouping	S04, S61 and S106	3	12
Correlation Analysis	S106 and S143	2	8
Structural equation modeling	S473	1	4
Multivariate Variance Analysis	S474	1	4
Statistical Learning	S46	1	4
No techniques presented	S00, S63, S99, S126, S148, S233, S326, S336, S349, S377, S381	11	44

Regression analysis is a statistical technique applied in classification analysis and used to verify the existence of a relationship between dependent variable and independent variables. Clustering is a statistical method that aims to explore the data generating hypotheses and assisting in prediction processes. In this SLR, we identified that 32.25% (10 studies) used Regression Analysis techniques, 12% (3 studies) reported using statistical data clustering techniques and 8% (2 studies) used Correlation Analysis. Structural Equation Modeling, Multivariate Variance Analysis and Learning Statistics Method were used in 1 study (4%) respectively. 11 other studies (44%) do not report using any statistical technique. Articles selected from String 3 (P1, P11, P125, P717, P729 and P744) do not present statistical analysis techniques.

Studies S326, S349 and S377 propose a model for predicting student performance in distance education. In these studies, the use of statistical techniques to analyze the data was not mentioned. They make use of a set of Machine Learning algorithms to perform the prediction.

Among the collected studies, a significant portion 32.25% (10 studies) cite using specific training algorithms and data testing in their experiments. Table 16 presents the distribution of these studies among the different algorithms identified.

The studies presented in Table 16 use Machine Learning algorithms to perform academic data evaluation, especially in predicting student success or possible dropout.

Table 16
Identified Algorithms for Evaluating EMS Student Data

Technique	Studies	Freq.	%
Support Vector Machine	S46, S349	2	22.22
N-Gram (Cluster)	S183	1	11.11
J48 Decision tree	S208	1	11.11
C4.5 (Decision tree)	S354, S377	2	22.22
JRip	S208	1	11.11
Logistic Regression	S208, S326, S421	3	33.33
Multilayer Perceptron (MLPs)	S208, S326	2	22.22
Naive Bayes	S208, S326, S377	3	33.33
Minimal Sequential Optimization (SMO)	S208	1	11.11
K-means	S04	1	11.11
RandomTree	S326	1	11.11
NNge	S326	1	11.11
GP-ICRM	S326	1	11.11
Radial Base Function	S349	1	11.11
General Bayesian Network (GBN)	S354	1	11.11
WINNOW	S377	1	11.11
Nearby Neighbor's Algorithm (1-Nearest Neighbor – 1NN)	S377	1	11.11

4.4. *Performance Indicators*

RQ3 focuses on the aspects of interaction mechanisms and performance indicators in the EMS. We seek to collect all the indicators used in the selected studies to identify or measure academic performance.

RQ3 – What elements of LMS are used as learning indicators?

This question aims to identify the indicators, available in the LMS, that can express the student's academic performance. The indicators were classified according to the taxonomy suggested by Rienties and Toetenel (2016). The taxonomy defines a classification of Activities in an LMS through verbs representing user actions. These activities are classified into: Assimilative, Communicative, Productive, Interactive, and Experimental Activities (Table 17).

Assimilative Activities include actions such as reading, listening, thinking about, accessing information and content. Communicative Activities are associated with discussion modules that are related between LMS users and content (tutors, teachers and other students). Productive Activities relate actions in the active construction of artifact, such as producing, building, making, contributing and completing. Interactive Activities comprise learning application actions such as exploring, experimenting, interacting with content, enhancing and modeling. As for Evaluative Activities, these relate to types of evaluations (formative, summative and self-assessment). Finally, the experimental activities that address the learning actions applied in the real world are included.

Table 17 shows that a study may be represented in more than one classification because there are several learning indicators in the environment. The use of one indicator does not exclude another; on the contrary, the combination may provide more powerful data on student learning experiences and favor the monitoring of their academic performance.

The data presented in Table 17 highlight three Activity indicators: Productive, Interactive, and Evaluative, which are represented in more than 50% of the sample. The

Table 17
Indicators of use of activities and interactions in the Learning Management System

Description	Studies	Freq.	%
Productive Activities	S00, S04, S09, S63, S66, S143, S148, S183, S208, S233, S319, S326, S336, S349, S473, S474	16	64
Interactive Activities	S00, S04, S09, S61, S66, S99, S143, S148, S183, S233, S319, S336, S354, S421, S473, S474	16	64
Evaluative Activities	S00, S04 S09, S46, S61, S63, S106, S43, S208, S233, S319, S336, S377, S381, S473	15	60
Assimilative Activities	S00, S09, S46, S61, S63, S66, S148, S143, S349, S354, S473	11	44
Communicative Activities	S00, S126, S143, S208, S233, S326, S336, S354, S421, S473, S474	11	44

Productive and Interactive Activities category was cited in more than 64% (16 studies) among all selected articles (31). Second is the Evaluative Activities category with 60% (15 studies), this category determines the students' exams and evaluative activities. In third place, the Interactive and Communicative Activities category was also well represented (44%, 11 studies).

The result is interesting because it shows that most of the studies cite the interactions of Productive and Interactive activities as a strong indicator in monitoring student performance, especially when these indicators relate to other types (Evaluative and Communicative). This is noticeable because an article is represented in more than one category, such as articles S473 and S474. The Communicative Activities category represents less than 50% of the selected studies. This category is related to discussion modules and messages about content, people and the learning environment.

4.5. *Taxonomy of Educational Objectives*

In order to answer the RQ5 question, we sought to identify the taxonomies developed, referenced and used in the e-Learning teaching modality.

RQ5 – What are learning objective classification hierarchical structures used to monitoring student academic progress?

This question aims to find the taxonomic structures of educational objectives used in the teaching process to monitor student performance. We classify the studies according to the type of taxonomy cited in the paper. Some authors cite more than one taxonomy in their studies, see Table 18.

The data in Table 18 show the taxonomic structures identified among the filtered studies. We identified 14 studies (45.16%) that report using some taxonomy in the research context. Among the studies analyzed, Bloom's Taxonomy is the most referenced (29.03%, 9 studies), followed by Bloom's Revised Taxonomy (6.45%, 2 studies). The other taxonomies (SOLO Taxonomy, Learning Design Activities Taxonomy and Learning Strategies and Techniques Taxonomy) were identified in at least one study (3.22%).

Table 18
Learning Taxonomy identified in the selected studies

Taxonomy	Studies	Freq.	%
Bloom Taxonomy	S00, S63, S126, S381, P1, P125, P717, P729, P744	9	29.03
Bloom's Revised Taxonomy	S00, P11	2	6.45
SOLO Taxonomy	S63	1	3.22
Learning Design Activities Taxonomy	S09	1	3.22
Taxonomy of learning strategies and techniques	S99	1	3.22

5. Discussion

The data identified in this SLR show that 56% (19 studies) focus on predicting the success or failure of student academic progress; 36% (9 studies) aim to evaluate academic performance; 4% (1 study) has a competence-based recommendation system and another study (4%, 1 study) aims to classify evaluative questions based on Bloom's Taxonomy. Table 19 presents the classification of the study objectives in relation to the use of AT techniques.

Academic performance prediction systems perform student analysis predicting whether students will fail or succeed in their endeavors. Machine Learning (ML) techniques and algorithms are commonly applied for this purpose since some of these algorithms use past data (test basis) to analyze patterns to predict the student's situation. Prediction systems typically induce educators to make early pedagogical decisions and strategies to improve student learning conditions and the dropout scenario.

As for the works that present some student assessment method in e-Learning environments, the use of some statistical techniques was identifying. Table 20 lists the techniques identified.

According to Larose (2005), the data mining process is classified by the ability to perform specific tasks, among them stand out: Regression, Prediction, Clustering, and Association. Larose (2005) describes that each task uses specific statistical methods classified according to the desired functionality (Descriptive or Prognostic Analysis) and the tasks they perform. Table 21 presents methods that can be used for each task or functionality found.

Table 19

Classification of objectives in identified studies

Objective	Studies	Freq.	%
Prediction of academic performance	S09, S46, S106, S143, S148 S183, S208, S319, S326, S336, S349, S354, S377, S421	14	56
Evaluate academic performance	S00, S04, S61, S66, S99, S233, S126, S473, S474	9	36
Classification of evaluation questions	S63	1	4
Competency-based recommendation system	S381	1	4

Table 20

Classification of techniques used in performance evaluation

Technique	Studies	Freq.	%
Data grouping	S04, S61	2	22.22
Regression Analysis	S61, S66, S126	3	33.33
Structural equation modeling	S473	1	11.11
Multivariate Variance Analysis	S474	1	11.11
Not specified	S00, S99, S233	3	33.33

Table 21

Set of Methods Used in Each Data Mining Task

Assignment	Methods
Grouping	• *Partitioning Methods based on the grid* • *Model-based clustering methods – statistical approach and neural networks* • *Analysis of outliers* • *Hierarchical Methods* • *Density based methods*
Regression	• *Linear Regression* • *Multiple Regression* • *Logistic Regression* • *Poisson Regression* • *Nonlinear Regression*

The methods linked to cluster analysis aim to detect the existence of different groups within a given data set and, if so, to determine these groups. Clustering attempts to identify a finite set of categories or groups to which each record (population element) can be mapped. Categories can be disjoint (separate) or overlapping (non-disjoint) and can sometimes be arranged in trees.

Regression methods are used when the record is identified by a numerical value rather than a categorical value. Thus, the value of a given variable can be estimated by analyzing the values of the others. According to Harrison (1998), regression is used to define a value of a given unknown continuous variable. Fayyad *et al.* (1996) define regression as a function that maps a data item to an estimated actual prediction variable.

Regarding the use of Ontologies in the context of LA and LMS, we identified 9 papers (S63, S99, S381, P1, P11, P125, P717, P729, and P744) that mention using ontologies and taxonomic learning structures linked to an LMS. Among these works, only two (S99 and S381) aim to analyze and improve student performance. In study S381, the authors present a theoretical architecture using a network of ontologies to identify and evaluate student competences and recommend learning objects.

Study S99 presents a conceptual architecture for detecting and analyzing cognitive and metacognitive learning activities in personal learning environments. The learning ontology used consists of cognitive and metacognitive learning activities that describe typical learning activities. The pedagogical perspective of this approach focuses on the reflection and awareness aspects of the learning process.

Study S63 aims to verify at which level the examination questions formulated by educators fit into Bloom's taxonomic structure. The authors use Natural Language Processing techniques to extract the verbs, then verify the similarity of the extracted verbs by querying the WordNet base (Top-level Ontology), and finally classify the issue according to the taxonomic level.

The remaining 6 studies (P1, P11, P125, P717, P729, and P744) present ontological structures applied in the educational context. Most of these studies focus on learning sequencing and selecting Virtual Learning Objects (VLE) related to the student's cognitive level. A taxonomic structure of the Learning Objectives is used from the perspective of measuring whether the VLE level corresponds with the student's learning sequencing.

Regarding the Taxonomies of Educational Objectives cited in the selected studies, we identified that Bloom's Taxonomy is the most referenced and used. This taxonomy was pioneering, consolidated by educational theories and is used as a basis for the development of other taxonomic structures.

Regarding the indicators used to perform the academic performance evaluation, most studies use the Productive and Interactive Activities. As shown in Table 17, some studies use a set of indicators to enable more accurate performance monitoring. The evaluative activities are usually used because they have a record of the exams performed in the LMS. However, the coordinated use of performance indicators will provide a better representation of student academic performance.

As student learning takes place through their interactions with the LMS, in study S00 the authors present the use of the *xAPI* framework (*Experience API*) to collect LMS data and perform cognitive processing of the student. *xAPI* is a formal specification for educational technologies that enables the collection of data about the wide variety of experiences a person has on a learning platform. The *xAPI* specification makes it possible to collect student actions in the LMS through statements represented by a triple in RDF consisting of: Actor – Verb – Object, as an example: "John watched Pipeline Video" or "*John completed the Pipeline Activity with note 8*". The *xAPI* framework has a wide vocabulary of verbs[4] that represent students' various interactions with planned activities in the LMS.

The results of this review show that there is a possible research gap. Few works that use Computational Ontologies and LA techniques in a coordinated way guided by a Taxonomy Learning Objectives for the monitoring of academic performance were identified. There is a need for computational tools that help educators to evaluate educational data and to monitor academic progress consistently in e-Learning environments. A summary of the evidence and information mapping is presented in Table 22.

Table 22
Summary of evidence and mapping identified in the review

Searched Object	Most referenced element	Studies	%
Identified Taxonomy	Bloom Taxonomy	S00, S63, S126, S381, P1, P125, P717, P729, P744	29.03
Learning Environment	MOOC	S04, S126, S183, S326, S354, P744	19.35
Ontology	Learning Ontology	S63, S99, S381, P1, P11, P125, P717, P729, P744	29.03
Indicator	Productive Activities	S00, S04, S09, S63, S66, S143, S148, S183, S208, S233, S319, S326, S336, S349, S473, S474	51.61
	Interactive Activities	S00, S04, S09, S61, S66, S99, S143, S148, S183, S233, S319, S336, S354, S421, S473, S474	51.61
Methods	Regression Analysis	S61, S66, S126	9.67

[4] `http://xapi.vocab.pub/verbs/index.html`

The results of the SLR show that among the solutions identified for monitoring student academic progress, no work uses coordinated LA and Ontologies to track student academic performance guided by a Taxonomy of Educational Objectives.

The LA application methodology enables the collection of educational data, processing, analysis and presentation of student information. LA enhances the teaching and learning process as it can extract useful information about all interactions that occur in the LMS. LA is seen as another layer of analysis of existing educational data, as teachers cannot manually handle or process the volume of data produced in the environment. The learning management reports available in the LMS are not produced with a wealth of information regarding student academic performance.

Regarding the use of Computational Ontologies, we observed several works that use them to support software agents such as understanding the abstract concepts related to a specific domain, and assisting in the processing of educational data. The use of ontologies in the educational context has been adopted by several applications specific to Distance Education.

In the context of this research, a parameterizable educational software architecture is prospected to monitor students' academic performance. The architecture will enable the educator to choose a Taxonomy of Educational Objectives, formalized by ontologies, which will parameterize the entire data processing and analysis module by performing a more efficient evaluation of educational data.

6. Threats to Validity

This section describes potential threats to the validity of this SLR and concerns about future reproductions. The section was organized according to Dermeval *et al.* (2017) and classification presented by Wohlin *et al.* (2012) that defines threats in the Internal, External, Construction and Completion categories.

Threats to internal validity are features that aim to mitigate systematic errors within the circumstances of the review. During study selection and data extraction, some subjective decisions may have occurred due to the lack of a clear description and adequate results from the primary studies. This situation determines a detection bias, as some results may be coded or misinterpreted, which in many cases makes it impossible to identify evidence. This scenario makes it difficult to apply the inclusion and exclusion criteria, as well as extracting data. To minimize selection and extraction errors, the process was performed iteratively and extraction collaboratively by the authors. Thus, we seek to mitigate the threats regarding the understanding of a particular study.

As for threats to external validity, these are related to the possibility of generalizing the review results and the extent to which the identified primary studies are representative for the object of the review. To mitigate external threats, the search process (presented in Sections 3.2.1 and 3.2.2) was defined after validation and consensus by the authors. Thus, the coverage and representativeness of the retrieved studies including automatic database searching were tested.

The main constructs for this review are three concepts, “Learning Analytics”, “Ontologies” and “Educational Objectives”.

For the first concept, we use the terms “Learning Analytics” and “Educational Data Mining”, because LA comes from Educational Data Mining and we needed more studies related to the research object. We sought to ensure that all selected studies are related to the Learning Analytics OR Educational Data Mining approach.

The second concept “Ontology” when inserted into the search expression using the logical operator “AND”, did not present results of studies related to the automatic execution of the expression in search engines. We decided to omit the term “Ontology” from String 2, as its omission does not result in the exclusion of works related to this term, since the search expression remains generic, and to create a String (String 3) to cover studies related to this term Ontologies.

As for the third concept, we use the term “Learning Objectives”, its synonyms and related terms (“educational objective” OR “instructional objective” OR “learning objective” OR “cognitive objective” OR “cognitive process” OR “cognitive learning” OR “ educational theory “OR” learning theory “), to ensure high coverage of potentially relevant studies from an automatic search in digital libraries.

Threats to the conclusion validity relate to issues that affect the ability to draw the correct conclusion about the relationships between treatment and review outcome, for example, it is possible that some excluded studies should have been included. In order to mitigate this threat, inclusion and exclusion criteria were carefully developed and discussed among the authors. These criteria help to reduce personal bias and guide the study selection process. Candidate exclusion studies were discussed among researchers in order to reach a consensus and determine greater representativeness.

7. Conclusions

In this work, a Systematic Literature Review was carried out to investigate the coordinated use of LA and Computational Ontologies guided by a Taxonomy of Educational Objectives aiming to monitor the academic performance of students in Distance Education. Our purpose was to improve the understanding of the use of analytical techniques that can assist in the processing of educational data and the deepening of methods that allow students to monitor their academic performance in order to promote learning.

Thirty-one studies out of a total of 1230 papers were selected to provide information to answer one main research and five other secondary questions. The results show a significant increase in the number of studies, revealing a growing interest in this area of research, and also identify a trend: in recommending Virtual Learning Objects and in predicting the success or failure of student learning. A relevant research gap was noted, as we did not identify a significant number of articles using Computational Ontologies and LA in a coordinated manner, and more specifically, no work was identified using Taxonomies of Educational Objectives to evaluate fulfillment of the educational objectives and performance of students. In addition, the papers retrieved in this SLR also shows that there is a need for tools that help educators to consistently monitor the academic progress of the distance learning students.

The results also suggest that:

1) Among the five types of activities that direct the performance indicators on the learning platforms, three activities (productive, interactive, and evaluative) were much more targeted for academic performance observation. We believe that an analysis based on all indicators can provide more consistent information about a student's academic progress.
2) Different learning platforms have been identified, consequently, the standardization and modelling of the collected data to remove any noise is necessary. Standardization of data is necessary to ensure the quality of the results produced by the application of the Learning Analytics.
3) Generally educators use a Taxonomy of the Learning Objectives to support their pedagogical plan and in this research, some taxonomies were identified. Thus, the formalization of these taxonomies will enable us to parameterize analysis tools, which will understand the taxonomy used by the educator. The analysis tools will help to understand the structure of the classification of the educational objectives that the educators use, besides make an analysis and inferences about the students learning experiences.
4) A diversity of algorithms for classifying and clustering educational data was identified. It is interesting to carry out an analysis of these algorithms by applying them in the educational dataset from learning platforms, in order to observe the precision and performance of the results of each algorithm.

The results presented in this systematic review can be useful for researchers in the areas of Data Mining, Analytics, Big Data, and Ontologies applied in an educational context, since it gathers evidence from primary studies. Such evidence indicates the use of a more adaptive and personalized learning environment, and the better use of pedagogies to enhance teaching/learning.

As future work, the development of a parameterizable architecture model using LA and Computational Ontologies to assist educators in the process of student assessment in Distance Education is expected. The parameterization of the model will allow the educator to adjust the architecture according to the selected taxonomy and will enable a more consistent assessment of student learning experiences.

Acknowledgments

This work was supported by the Institutional Qualification Program (PIQ) of the Federal Institute of Sertão Pernambucano (IFSertão-PE, Campus Petrolina).

References

Achimugu, P., Selamat, A.; Ibrahim, R., Mahrin, M. N. (2014) A systematic literature review of software requirements prioritization research. *Information and Software Technology*. 56(6), 568–585. DOI: 10.1016/j.infsof.2014.02.001

Alraimi, K. M., Zo, H., Ciganek, A. P. (2015). Understanding the MOOCs continuance: the role of openness and reputation. *Journal of Computers & Education*, 80, 28e38. DOI: 10.1016/j.compedu.2014.08.006.

Al-Yahya, M.; George, R.; Alfaries, A. (2015) Ontologies in E-Learning: Review of the Literature. *International Journal of Software Engineering and Its Applications*, 9(2), 67–84. DOI: 10.14257/ijseia.2015.9.2.07

Amorim, R. R., Lama, M., Sánchez, E., Riera, A., & Vila, X. A. (2006) A Learning Design Ontology based on the IMS Specification. *Educational Technology & Society*, 9(1), 38–57. ISSN-1436-4522.

Anderson, L., Krathwohl, D. (2001) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Publisher: New York, ISBN: 080131903X Edition: 1st.

Anseeuw, J., Verstichel, S., Ongenae, F., Lagatie, R., Venant, S., Turck, F. (2016) An Ontology-enabled Context-aware Learning Record Store Compatible with the Experience API. In Proceedings of the 8th International Joint Conference on Knowledge Discovery, *Knowledge Engineering and Knowledge Management* (IC3K 2016) – Volume 2: KEOD, pages 88–95. DOI: 10.5220/0006049000880095.

Archer, E., Chetty, B., Prinsloo, P. (2014) Benchmarking the habits and behaviors of successful students: A case study of academic-business collaboration. *The International Review of Research in Open and Distance Learning*, 15(1). http://www.irrodl.org/index.php/irrodl/article/download/1617/2810/0

Biggs, J., Collis, K. (1982) *Evaluating the Quality of Learning: The SOLO Taxonomy*. New York: Academic Press. ISBN 9780120975525. DOI: 10.1016/B978-0-12-097552-5.50005-3

Bloom, B.S., Engelhart, M.D., Furst, E., Hill, W., Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: David McKay Company.

Bourdeau, J., Mizoguchi, R., Hayashi, Y., Psyche, V., Nkambou, R. (2007) When the Domain of the Ontology is Education, 4ª Conference on *Intelligent, Interactive Learning Objects Repository Networks*, Montreal, in: Proc. of the 4th Conf. on Intelligent, (I2LOR'07). pp. 1–9.

Buchinger, D., Cavalcanti, G.A.D.S., Hounsell, M.D.S. (2014) Academic search mechanisms: a quantitative analysis. *Brazilian Journal of Applied Computing*, 6(1), 108–120. DOI: 10.5335/rbca.2014.3452

Campbell, J. P., DeBlois, P. B., Oblinger, D.G. (2007). *Academic Analytics: A New Tool for a New Era. Educause*. https://www.wsac.wa.gov/sites/default/files/2014.ptw.(63).pdf

Costa, L.A., Salvador, L.N. (2015). In-person and virtual learning environment integrated with ubiquitous computing: A systematic mapping of the literature. In: *XX International Congress of Educational Informatics*. Santiago, University of Chile, v. 11, p. 211–220.

Costa, L.A., Salvador, L.N., Amorim, R.R. (2018) Evaluation of Academic Performance Based on Learning Analytics and Ontology: A Systematic Mapping Study. *IEEE Frontiers in Education – FIE*, San Jose, USA, pp. 1–5. DOI: 10.1109/FIE.2018.8658936.

Dermeval, D., Paiva, R., Bittencourt, I.I., Vassileva, J., Borges, D. (2017) Authoring Tools for Designing Intelligent Tutoring Systems: A Systematic Review of the Literature. *International Journal Artificial Intelligence in Education*. 28, 336–384. DOI: 10.1007/s40593-017-0157-9.

Dyba, T., Dingsoyr, T. (2008) Strength of evidence in systematic reviews in software engineering. *Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, n. 7465, p. 178–187. DOI: 978-1-59593-971-5/08/10.

Easterbrook, S., Singer, J., Storey, M., Damian, D. (2008) Selecting empirical methods for software engineering research. In: Shull F., Singer J., Sjberg D. (eds) *Guide to Advanced Empirical Software Engineering*, Springer London, pp 285–311. DOI: 10.1007/978-1-84800-044-5_11.

Educause. (2013). *Seven Things you Should Know about MOOCs II*. Educause learning initiative. https://library.educause.edu/-/media/files/library/2013/6/eli7097-pdf.pdf

Elias, T. (2011) *Learning Analytics: The Definitions, Processes, and the Potential*. https://pdfs.semanticscholar.org/732e/452659685fe3950b0e515a28ce89d9c5592a.pdf

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). From data mining to knowledge discovery: An overview. In: *Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Press, England, p.1–34.

Fernández-Delgado, M., Mucientes, M., Lama, M., Vázquez-Barreiros, B. (2014). Learning Analytics for the Prediction of the Educational Objectives Achievement. 44th *Annual Frontiers in Education Conference (IEEE Frontiers in Education 2014)*, pp. 2500–2503, Madrid, Spain. DOI: 10.1109/FIE.2014.7044402.

Freire, Paulo (2000). *Pedagogy of Autonomy: Knowledge Necessary for Educational Practice*. (In Portuguese). 15 ed. São Paulo: Peace and Earth. ISBN 8521902433.

Ghani, I., Yasin, I. (2013). Software security engineering in extreme programming methodology: a systematic literature. *Science International (Lahore)*, 25(2), 215–221. ISSN 1013-5316.

Goodyear, P., Retalis, S. (2010). Technology-Enhanced Learning, Design Patterns and Pattern Languages. *Sense Publishers*. Volume 2.
`https://www.sensepublishers.com/media/1037-technology-enhanced-learning.pdf`

Gruber, T.R. (1993). *A Translation Approach to PorTable Ontology Specifications. Knowledge Acquisition*, 5(2):199–220. `https://doi.org/10.1006/knac.1993.1008`

Harrison, T.H. (1998) *Intranet data warehouse: Ferramentas e Técnicas para Utilização do Data Warehouse na Intranet*. (in Portuguese) São Paulo: Bekerley Brazil.

Haydt, R.C.C. (2011). *General Didactics course*. (in Portuguese) 1ª Ed. – São Paulo: Ática.

Higgins, J., Thomas, J. (2008). *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.0.0, The Cochrane Collaboration. `http://training.cochrane.org/handbook`

Hrastinski, S. (2009). A theory of online learning as online participation. *Journal of Computers & Education*, 52(1), 78–82. Elsevier. DOI: 10.1016/j.compedu.2008.06.009.

Iglesias-Pradas, S., Ruiz-de-Azcárate, C., Agudo-Peregrina, A.F. (2015) Assessing the suitability of student interactions from Moodle data logs as predictors of cross-curricular competencies. *Journal Computers in Human Behavior*, Vol. 47, p. 81–89. Elsevier. DOI: 10.1016/j.chb.2014.09.065.

Khalil, M., Ebner, M. (2015). Learning analytics: principles and constraints. In: Proceedings of world conference on educational multimedia, hypermedia and telecommunications, pp 1326–1336. *Proceedings of EdMedia – World Conference on Educational Media and Technology* (pp. 1789–1799). Montreal, Quebec, Canada: Association for the Advancement of Computing in Education (AACE).
`https://www.learntechlib.org/primary/p/151455/`

Kitchenham, B., Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Technical Report EBSE-2007-01, Keele University and University of Durham.

Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review, *Information and Software Technology*, Volume 51, Issue 1, 2009, Pages 7–15, DOI: 10.1016/j.infsof.2008.09.009.

Korchi, A.; Abdellah, M. B.; (2015). A modeling learner approach in a computing environment for human learning based on ontology, *International Journal of Computer Engineering & Technology*.

Lang, C., Siemens, G., Wise, A., Gasevic, D. (2017). Handbook of Learning Analytics – First Edition. Solar – *Society for Learning Analytics Research*. DOI: 10.18608/hla17.

LaPES (2013). *Start – state of the art through systematic review tool*.
`http://lapes.dc.ufscar.br/tools/start_tool`

Larose, D. T. (2005) Discovering Knowledge in Data: An Introduction to Data Mining. *John Wiley and Sons*. Chichester; ISBN 0. 471 66657 2. USA.

Li, X., Wang, T., Wang, H. (2017). Exploring N-gram Features in Clickstream Data for MOOC Learning Achievement Prediction. In: Bao Z., Trajcevski G., Chang L., Hua W. (eds) *Database Systems for Advanced Applications*. DASFAA 2017. Lecture Notes in Computer Science, vol 10179. Springer, Cham. DOI: 10.1007/978-3-319-55705-2_26.

Lima, R.W., Fialho, S.V. (2011). Content Map and Dependency Map: tools for planning based on educational objectives. In: *Exact and Technological Magazine*. RETEC, v.2, p.10.

Luckesi, C.C. (2011). *Learning Assessment – Component of the pedagogical*. (in Portuguese) ISBN: 9788524916571. Cortez Publisher.

Macfadyen, L., Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology and Society*, 15(3), 149–163. Access in: February 2019. Available in: `https://www.jstor.org/stable/jeductechsoci.15.3.149`

Mahdavi-Hezavehi, S., Galster, M., Avgeriou, P. (2013). Variability in quality attributes of service-based software systems: A systematic literature review. Information and Software Technology 55(2):320-343, *special Section: Component-Based Software Engineering (CBSE)*. DOI: 10.1016/j.infsof.2012.08.010.

Margaryan, A., Bianco, M., Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers & Education*, 80, 77e83. DOI: 10.1016/j.compedu.2014.08.005.

Muñoz, A., Lasheras, J., Capel, A., Cantabella, M., Caballero, A. (2015). OntoSakai: On the optimization of a Learning Management System using semantics and user profiling. *Expert Systems with Applications*, v. 42, n. 15–16, p. 5995–6007. DOI: 10.1016/j.eswa.2015.04.019.

Pelissoni, A.M.S. (2009). *Educational Objectives and Learning Assessment*. (in Portuguese) Teaching Academic Production Yearbook. Vol. III, No. 5. Anhanguera Educational, São Paulo.

Psyche, V., Bourdeau, J., Nkambou, R., Mizoguchi, R., (2005). Making Learning Design Standards Work with an Ontology of Educational Theories, in: in Proc. of the 12th Artificial Intelligence in Education (AIED2005), Amsterdan, The Netherlands, pp. 539–546.

Quinn, S.; Bond, R.; Nugent, C.; (2017). Ontological modelling and rule-based reasoning for the provision of personalized patient education, *Expert Systems*.

Rienties, B., Toetenel, L. (2016). The impact of 151 learning designs on student satisfaction and performance: social learning (analytics) matters. *In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)*. ACM, New York, NY, USA, 339–343. DOI: 10.1145/2883851.2883875.

Sampaio, R.F., Mancini, M.C. (2007). Systematic review studies: a guide for judicious synthesis of scientific evidence. *Brazillian Magazine*. vol.11 no.1 Sao Carlos. DOI: 10.1590/S1413-35552007000100013.

Sanches, L.P., Costa, L.A., Salvador, L.N., Souza, M.V. (2017). Automatic semantic annotation of Digital Learning Objects: A systematic literature mapping. *Brazilian Symposium on Informatics in Education-SBIE*. DOI=10.5753/cbie.sbie.2017.31.

Siemens, G., Long, P. (2011). *Penetrating the fog: Analytics in learning and education*. EDUCAUSE Review, 46(5), 31–40.
`https://er.educause.edu/~/media/files/article-downloads/erm1151.pdf%20`

Society for Learning Analytics Research. (2011). Open Learning Analytics: an integrated & modularized platform.
`https://solaresearch.org/wp-content/uploads/2011/12/OpenLearningAnalytics.pdf`

Taraghi, B., Ebner, M., Saranti, A., Schön, M. (2014). On using markov chain to evidence the learning structures and difficulty levels of one digit multiplication. *In Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK'14)* (pp. 68–72). New York, USA:ACM. DOI: 10.1145/2567574.2567614.

Vesin, B.; Ivanović, M.; Klašnja-Milićević, A.; Budimac, Z.; (2012). Protus 2.0: Ontology-based semantic recommendation in programming tutoring system, *Expert Systems with Applications*. 39, 15 (November 2012), 12229–12246. DOI: 10.1016/j.eswa.2012.04.052.

Villagra-Arnedo, C.J., Gallego-Duran, F.J., Llorens-Largo, F., Compañ-Rosique, P., Satorre-Cuerda, R., Molina-Carmona, R. (2017). Improving the expressiveness of black-box models for predicting student performance. *Journal Computers in Human Behavio*r, Vol. 72, p. 621–631. Elsevier. DOI: 10.1016/j.chb.2016.09.001.

Wohlin, C., Runeson, P., Host, M., Ohlsson, M., Regnell, B., Wessl´en, A. (2012). Experimentation in software engineering. *Springer Science and Business Media*. DOI: 10.1007/978-3-642-29044-2.

Yago, H., Clemente, J., Rodriguez, D., Fernandez-de-Cordoba, P. (2018). ON-SMMILE: Ontology Network-based. *Data & Knowledge Engineering*, Elsevier. Volume 115, Pages 48–67, ISSN 0169-023X, DOI: 10.1016/j.datak.2018.02.002.

Zhang, C., Zhang, J. (2010). InForCE: Forum data crawling with information extraction. *4th International Universal Communication Symposium*, Beijing, 2010, pp. 367–373. DOI: 10.1109/IUCS.2010.5666252.

Zimmerman, B.J. (2002). Becoming a Self-Regulated Learner: An Overview, *Theory Into Practice*, 41:2, 64–70, DOI: 10.1207/s15430421tip4102_2.

L.A. Costa, graduated in Computer Science from the Faculty of Applied and Social Sciences of Petrolina-FACAPE (2004), specialist in Cryptography and Network Security from the Federal Fluminense University – UFF (2006), master in Computer Science at the Federal University of Pernambuco – UFPE (2014), and student of Ph.D. in Computer Science at Federal University of Bahia – UFBA. Currently work Effective Professor of the Informatics Collegiate at the Federal Institute of Sertão Pernambucano, Campus Petrolina, developing research in applied computing in the education.

L.M. Pereira Sanches, graduated in Computer Engineering at the University of Ribeirão Preto (UNAERP) with a sandwich period at the Beijing Institute of Technology (BIT) funded Higher Education Personnel Improvement Coordination (CAPES) as part of the Science Without Borders Program, Master in Computer Science at Federal University of Bahia (UFBA) with a grant from the Bahia State Research Support Foundation (FAPESB) and the Coordination for the Improvement of Higher Education Personnel. Develops research projects related to Semantic Web and distance education.

R.J. Rocha Amorim, Post Doctorate, Doctorate and Graduation in Computer Science. He is currently coordinating a research project at the National Institute of Science and Technology for Software Engineering (INES), working in the area of Computer Science with emphasis on Artificial Intelligence, Knowledge Representation and Software Engineering. Also works in the area of Informatics in Education with emphasis on Educational Ontology, Patterns of Educational Metadata, Learning Objects, Learning Design and Learning Analytics.

L. do Nascimento Salvador, graduated in Data Processing from the Federal University of Bahia (1990), master's degree in Electrical Engineering from the University of São Paulo (1995) and Ph.D. in Electrical Engineering from the University of São Paulo (2001). She is currently an associate professor at the Department of Computer Science at the Federal University of Bahia. Has experience in Informatics in Education, Computer Education, and Ontology Engineering, acting on the following subjects: data integration, semantic annotation, problem-based learning and information retrieval.

M.V. dos Santos e Souza, graduated in Mathematics from the Federal University of Bahia (2009), Graduated in COMPUTER SCIENCE from Salvador University (2008), Master in Computer Science from the Pontifical Catholic University of Rio Grande do Sul, PUCRS (2012), and Ph.D. in Computer Science from the Federal University of Rio Grande do Sul, UFRGS (2016). Has experience in Computer Science, focusing on Logic and Program Semantics, acting on the following subjects: PLN, formal programming language semantics, free software.

APPENDIX A

Below are two tables with the list selected studies during the process of the collection in the systematic literature review. The tables are classified according to the collection sequence in the Digital Libraries. There are two IDs: P and S. The ID ‘P’ represents papers collected with the execution of the String 3 (A1), and the ID ‘S’ refers to the studies recovered using the String 5 (A2).

A1. String 3 Results

Table A1

String 3 Results

#	ID	Reference
1	*P1*	Ghannem, A. (2014) Characterization of serious games guided by the educational objectives. In Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM ‘14). ACM, New York, NY, USA, 227–233. DOI: 10.1145/2669711.2669904.
2	*P11*	Ramesh, R., Sasikumar, M., Iyer, S. (2016) Integrating the Learning Objectives and Syllabus into a Domain Ontology for Data Structures Course. In Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE ‘16). ACM, New York, NY, USA, 266–271. DOI:10.1145/2899415.2899453.
3	*P125*	Peters, R., Broekens, J., Neerincx, M.A. (2017) Guidelines for Tree-based Collaborative Goal Setting. In Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI ‘17). ACM, New York, NY, USA, 401–405. DOI:10.1145/3025171.3025188.
4	*P717*	Guettat, B., Farhat, R. (2015) An approach to assist learners to identify their learning objectives in personal learning environment (PLE), 2015 5th International Conference on Information & Communication Technology and Accessibility (ICTA), Marrakech, 2015, pp. 1–6. DOI: 10.1109/ICTA.2015.7426934.
5	*P729*	Fulantelli, G., Taibi, D., Arrigo, M. (2015) A framework to support educational decision making in mobile learning, Computers in Human Behavior, Volume 47, 2015, Pages 50–59, DOI: 10.1016/j.chb.2014.05.045.
6	*P744*	Antonelli, D., D’Addona, D.M., Maffei, A., Modrak, V., Putnik, G., Stadnicka, D., Stylios, C. (2019) Tiphys: An Open Networked Platform for Higher Education on Industry 4.0, Procedia CIRP, Volume 79, 2019, Pages 706–711, DOI: 10.1016/j.procir.2019.02.128.

A2. String 5 Results

Table A2

String 5 Results

#	ID	Reference
1	*S00*	Gibson, A., Kitto, K., Willis, J. (2014) A cognitive processing framework for learning analytics. International Conference on Learning Analytics and Knowledge. ACM, New York, 212–216. DOI: 10.1145/2567574.2567610.

Continued on next page

Table 2 – continued from previous page

#	ID	Reference
2	*S04*	Khosravi, H., Cooper, K.M.L. (2017) Using Learning Analytics to Investigate Patterns of Performance and Engagement in Large Classes. ACM Technical Symposium on Computer Science Education. New York, NY, 309–314. DOI: 10.1145/3017680.3017711.
3	*S09*	Rienties, B., Toetenel, L. (2016) The impact of 151 learning designs on student satisfaction and performance: social learning (analytics) matters. Sixth International Conference on Learning Analytics & Knowledge. ACM, New York, NY, 339–343. DOI: 10.1145/2883851.2883875.
4	*S46*	Fernández-Delgado, M., Mucientes, M., Vázquez-Barreiros, B., Lama, M., (2014), Learning analytics for the prediction of the educational objectives achievement, IEEE Frontiers in Education Conference (FIE), Madrid, pp. 1–4. DOI: 10.1109/FIE.2014.7044402.
5	*S61*	Strang, K. (2016) How student behavior and reflective learning impact grades in online business courses. Journal of Applied Research in Higher Education, 8 (3), pp. 390–410. DOI: 10.1108/JARHE-06-2015-0048.
6	*S63*	Jayakodi, K., Bandara, M., Meedeniya, D. (2016) An automatic classifier for exam questions with WordNet and Cosine similarity, Moratuwa Engineering Research Conference (MERCon), Moratuwa, 2016, pp. 12–17. DOI: 10.1109/MERCon.2016.7480108.
7	*S66*	Yamada, M., Okubo, F., Oi, M., Shimada, A., Kojima, K., Ogata, H. (2016) Learning analytics in ubiquitous learning environments: Self-regulated learning perspective. International Conference on Computers in Education. pp. 306–314. ISBN: 978-986847357-7.
8	*S99*	Nussbaumer, A., Scheffel, M., Niemann, K., Kravcik, M., Albert, D. (2012) Detecting and reflecting learning activities in personal learning environments. Workshop on Awareness and Reflection in Technology-Enhanced Learning. Pp. 125–131. Available in: `http://ceur-ws.org/Vol-931/paper10.pdf`.
9	*S106*	Pardo, A., Han, F., Ellis, R.A. (2017) Combining University student self-regulated learning indicators and engagement with online learning events to Predict Academic Performance. IEEE Transactions on Learning Technologies, pp. 82–92. DOI: 10.1109/TLT.2016.2639508.
10	*S126*	Wu, Y., Wu, W. (2018) A Learning Analytics System for Cognition Analysis in Online Learning Community. In: U L., Xie H. (eds) Web and Big Data. APWeb-WAIM 2018. Lecture Notes in Computer Science, vol 11268. Springer, Cham. DOI: 10.1007/978-3-030-01298-4_21.
11	*S143*	Tempelaar, D.T., Rienties, B., Giesbers, B., (2014) Computer assisted, formative assessment and Dispositional Learning Analytics in Learning Mathematics and Statistics. Computer Assisted Assessment. Research into E-Assessment. DOI: 10.1007/978-3-319-08657-6_7.
12	*S148*	Hlioui, F., Aloui, N., Gargouri, F., (2017) Automatic deductions of Learners – Profiling Rules Based on Behavioral Analysis. Computational Collective Intelligence. Springer, p. 233–243. DOI: 10.1007/978-3-319-67074-4_23.
13	*S183*	Li, X., Wang, T., Wang, H., (2017) Exploring N-gram Features in Clickstream Data for MOOC Learning Achievement Prediction. Database Systems for Advanced Applications. Springer, p. 328–339. DOI: 10.1007/978-3-319-55705-2_26.
14	*S208*	Kostopoulos, G., Lipitakis, A., Kotsiantis, S., Gravvanis, G. (2017) Predicting Student Performance in Distance Higher Education Using Active Learning. Engineering Applications of Neural Networks. Springer, p. 75–86. DOI: 10.1007/978-3-319-65172-9_7.
15	*S233*	Dalipi, F., Yayilgan, S.Y., Kastrati, Z. (2015) Enhancing the Learner's Performance Analysis Using SMEUS Semantic E-learning System and Business Intelligence Technologies. Learning and Collaboration Technologies. Springer, p. 208–217. DOI: 10.1007/978-3-319-20609-7_20.

Continued on next page

Table 2 – continued from previous page

#	ID	Reference
16	*S319*	Agudo-Peregrina, A.F., Iglesias-Pradas, S., Conde-González, Miguel, A., Hernández-García, A. (2014) Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. Computers in Human Behavior, Volume 31, Pages 542–550. DOI: 10.1016/j.chb.2013.05.031.
17	*S326*	Xing, W., Guo, R., Petakovic, E., Goggins, S. (2015) Participation-based student final performance prediction model through interpreTable Genetic Programming: Integrating learning analytics, educational data mining and theory. Computers in Human Behavior, Vol. 47, pp 168–181. DOI: 10.1016/j.chb.2014.09.034.
18	*S336*	Iglesias-Pradas, S., Ruiz-de-Azcárate, C., Agudo-Peregrina, A.F. (2015) Assessing the suitability of student interactions from Moodle data logs as predictors of cross-curricular competencies. Computers in Human Behavior, Vol 47, Pages 81–89. DOI: 10.1016/j.chb.2014.09.065.
19	*S349*	Villagrá-Arnedo, C.J., Gallego-Durán, F.J., Llorens-Largo, F., Compañ-Rosique, P., Satorre-Cuerda, R., Molina-Carmona, R. (2017) Improving the expressiveness of black-box models for predicting student performance, Computers in Human Behavior, Vol 72, p. 621–631. DOI: 10.1016/j.chb.2016.09.001.
20	*S354*	Xing, W., Chen, X., Stein, J., Marcinkowski, M. (2016) Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. Computers in Human Behavior, Vol 58, Pages 119–129. DOI: 10.1016/j.chb.2015.12.007.
21	*S377*	Kotsiantis, S., Patriarcheas, K., Xenos, M. (2010) A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. Knowledge-Based Systems, Vol 23, pp 529–535. DOI: 10.1016/j.knosys.2010.03.010.
22	*S381*	Yago, H., Clemente, J., Rodriguez, D., Fernandez-de-Cordoba, P. (2018) ON-SMMILE: Ontology Network-based Student Model for MultIple Learning Environments, Data & Knowledge Engineering, DOI: 10.1016/j.datak.2018.02.002.
23	*S421*	Zacharis, N.Z. (2015) A multivariate approach to predicting student outcomes in web-enabled blended learning courses, The Internet and Higher Education, Volume 27, 2015, Pages 44–53. DOI: 10.1016/j.iheduc.2015.05.002.
24	*S473*	Shorfuzzaman, M., Hossain, M.S., Nazir, A., Muhammad, G., Alamri, A. (2018) Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment, Computers in Human Behavior, Volume 92, 2019, Pages 578–588. DOI: 10.1016/j.chb.2018.07.002.
25	*S474*	Aljohani, N.R., Daud, A., Abbasi, R.A., Alowibdi, J.S., Basheri, M., Aslam, M.A. (2018) An integrated framework for course adapted student learning analytics dashboard, Computers in Human Behavior, Volume 92, 2019, Pages 679–690. DOI: 10.1016/j.chb.2018.03.035.