Data Collection and Visualization in Moodle: A Systematic Literature Review

Gildásio da Costa Teixeira
UFRN
Natal, Brazil
gildasio.teixeira@ufrn.br

Apuena Vieira Gomes
UFRN
Natal, Brazil
apuena.gomes@ufrn.br

Arnaud A. H. de Abreu
UFRN
Natal, Brazil
arnaud.holanda@ufrn.br

Francisca Soraya R. de M. Farias
UFRN
Natal, Brazil
soraya.farias@edu.natal.rn.gov.br

Carlos Franklin F. de Oliveira
UFRN
Natal, Brazil
franklin.ferreiraff@gmail.com

José Guilherme Villaça
UFRN
Natal, Brazil
guilhermevillaca.937@ufrn.edu.br

Luzia Daniele da S. Araújo
UFRN
Natal, Brazil
luziadaniele06@gmail.com

Micheli Gomes da Silva
UFRN
Natal, Brazil
micheli.silva@ufrn.br

ABSTRACT

This article aims to present, through a Systematic Literature Review (SLR), the results of an investigation into the forms, techniques, and applications available for use with Moodle for data collection and visualization. The study sought to answer the research questions: (1) How do studies present forms of data extraction from Moodle?; (2) What are the main dashboard/monitoring tools investigated for use with Moodle? The results provide valuable insights into how data visualization and activity monitoring tools can transform educational management in the context of Distance Education.

Author Keywords
Distance Education; Moodle; Learning management systems.

ACM Classification Keywords
CSS Applied computing - Education - Learning management systems

INTRODUCTION

In the current educational context, the increasing integration of digital technologies is redefining teaching methods and school governance. The adoption of platforms for institutional, administrative, and pedagogical monitoring is a promising initiative for optimizing educational and administrative practices (THIENY et al., 2023). However, it is necessary to delve deeper into how this adoption affects the activities of teachers, decision-making processes in school management, and, ultimately, the outcomes of the teaching-learning process in the context of distance education (DE).

According to data from the Open University of Brazil (UAB, 2024), the public education network currently has 141 Higher Education Institutions (HEIs), 913 active course offerings, and a total of 974 support centers. Table 01 shows how significant the numbers are for DE in public institutions at the federal level.

Level of Education	Active Offerings	Number of Active Students
Bachelor's Degree	63	8.904
Postgraduate (lato sensu)	215	32.739
Technologist	42	4.647

Licentiate	593	79.953
Total	913	126.241

Table 01 - Universidade Aberta do Brasil (2024)

The significance of these numbers alone highlights the need for monitoring that aligns with the scale and importance that DE in Brazil requires. After all, the words of MORAN (2002, p.2) remain valid and relevant: "DE is not 'fast food' where the student is served something ready-made. It is a practice that allows a balance between individual needs and skills and those of the group—in both face-to-face and virtual settings."

In this context, Moodle (Modular Object-Oriented Dynamic Learning Environment)[1] is one of the most widely used online learning platforms globally, being widely adopted by educational institutions at various levels to support distance and hybrid education. With the exponential growth in the use of Virtual Learning Environments (VLEs), data collection and visualization have become crucial elements in improving teaching effectiveness and the learning experience for students.

Data collection in Moodle encompasses various sources, including student interactions with content, completed activities, assessment results, and forum participation. When properly collected and analyzed, this data can provide valuable insights into student behavior, identify learning patterns, and assist in pedagogical decision-making. However, the challenge lies not only in data collection but also in how this data is presented and visualized for different stakeholders, such as teachers, administrators, and the students themselves (Porto; Dias; Battestin, 2023).

Data visualization plays a crucial role in transforming large volumes of raw data into understandable and actionable information. Effective visualization tools and techniques can help identify trends, monitor student progress, and personalize teaching according to individual needs. In the context of Moodle, data visualization can facilitate the understanding of engagement metrics, academic performance, and student retention, contributing to more informed and effective educational management (De Oliveira; Rodrigues; Maciel, 2022).

This study aims to conduct a systematic literature review on approaches to data collection and visualization in Moodle, highlighting the methodologies used, the challenges faced, and the solutions proposed. By reviewing studies published in the literature, we seek to provide a comprehensive view of current practices and identify areas that require further investigation.

The importance of this review lies in the growing need to improve educational practices through robust and efficient data analysis, as well as to form part of a doctoral project within a graduate program at a Federal University. With the widespread adoption of remote learning, the relevance of data collection and visualization in Moodle has become even more evident, as institutions seek ways to ensure teaching quality and support student success in a digital environment.

Therefore, this systematic review aims to answer the following research questions: How do studies present methods for extracting data from Moodle? What are the main dashboard/monitoring tools investigated for use with Moodle? It is expected that the results of this review will contribute to advancing knowledge in the field and provide practical guidelines for educators and VLE developers.

Method

Literature review aims to identify, analyze, and interpret existing and relevant evidence related to a research topic. The research topic can be better understood, and its relevance demonstrated by highlighting how the scientific community has approached the subject, thereby establishing its pertinence in the current context.

The research methodology for conducting this Literature Review was based on the guidelines and protocol proposed by Kitchenham and Charters (2007), and followed the following phases: protocol formulation, execution, data analysis, and interpretation. As supporting tools, the software Rayyan[2] and Zotero[3] were used. The following sections will present the research questions, search string, databases used, inclusion and exclusion criteria, quality assessment, and data extraction.

Research Questions

The research question for this Literature Review (LR) is derived from the PICOC strategy adapted from the work of BIOLCHINI et al. (2005). Table 2 presents the structuring of the elements that served to define the research questions.

Description	Description
Population	Users of the Moodle VLE in educational institutions
Intervention	Development of an information system that provides management, organization, analysis, and access to data within Moodle

[1] https://moodle.org/?lang=pt_br
[2] https://www.rayyan.ai/
[3] https://www.zotero.org/

Control	No need to compare with other technologies in the field, as the study does not require such an analysis.
Outcome	Overview of studies reported in the literature that discuss the use of plugins, external tools, and/or other solutions with Moodle.
Application Context	Educational institutions

Table 2: Description of the Research Criteria

Aiming to provide an overview of the studies conducted by the academic community on data extraction, monitoring, and visualization in Moodle[4], the following research questions were formulated.

- RQ1: How do studies present methods for extracting data from Moodle?
- RQ2: What are the main dashboard/monitoring tools investigated for use with Moodle?

Search Strategy/Selection Criteria

With the research questions defined during the review process, the next step was to develop the search strings to be used in search engines. Based on the previously mentioned PICOC strategy and the characteristics and specifics of the databases, three search strings in English were created. These English strings were applied in the search engines Scopus, IEEE, and ERIC. The search in Portuguese was conducted manually in the Periodicos Capes database. Table 3 displays the search strings used.

Database	Adapted String
SCOPUS	(TITLE-ABS-KEY ("moodle" OR "lms") OR TITLE-ABS-KEY ("modular object-oriented dynamic learning environment") AND TITLE-ABS-KEY (plugin AND (solution OR system OR application OR dashboard OR monitoring OR "data visualization"))) AND PUBYEAR > 2020 AND PUBYEAR < 2024
IEEE	("DOCUMENT TITLE":"MOODLE") OR ("ALL METADATA":"LMS") AND ("ALL METADATA":"MODULAR OBJECT-ORIENTED DYNAMIC LEARNING ENVIRONMENT") AND ("ALL METADATA":"PLUGIN")
ERIC	Moodle OR LMS AND "modular object-oriented dynamic learning environment" OR "plugin"
Periódicos Capes	Manual Search

Table 3: Search Strings for Each Database

Inclusion and exclusion criteria were defined to ensure that the selection of studies followed the guidelines outlined in Table 4.

Inclusion Criteria	Exclusion Criteria
CI1 - Primary articles	CE1 - Articles where Moodle was not used in an educational institution
CI2 - Full-text articles	CE2- Duplicate articles
CI3 - Published in English, Spanish, and Portuguese	CE3 - Gray literature
CI4 - published from 2020 onwards	CE4- Unavailable articles (download or access)
CI5 - Studies must address the research questions in at least one aspect	-

Table 4 - Inclusion and Exclusion Criteria

Selection

According to Nakagawa et al. (2017), the selection stage aims to ensure, based on the inclusion/exclusion criteria previously defined by the researchers, the relevance of the studies for addressing the research questions. As shown in Figure 1, the history and respective results of the protocol applied in this systematic review.

[4] This initial study becomes important as the platform to be developed can leverage tools already created for Moodle, taking advantage of what the community has to offer.

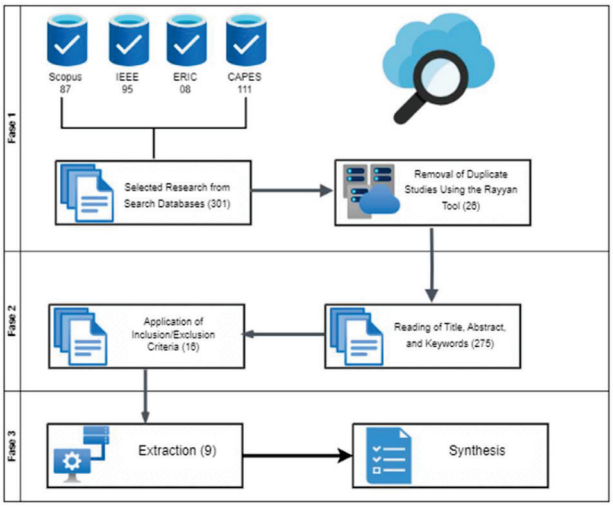

Figure 1 - Selection Process Flow

The selection of relevant studies was carried out in three stages. In Stage 01, the search strings were executed, returning the results presented in Table 2. Twenty-six duplicate articles were removed using the Rayyan tool. Subsequently, in Stage 02, the titles, abstracts, and keywords of the remaining 275 articles were reviewed, reducing the number of studies to 16, which are listed in the spreadsheet. These 16 articles were read in full and the inclusion and exclusion criteria (Table 3) were applied. Of the 7 excluded studies, 3 were due to exclusion criterion CE3, 3 due to CE2, and 1 due to CE1. As a result of the full reading of the 16 articles, as shown in Table 5, 9 articles were used for the extraction and synthesis of findings.

The results of these 09 articles were categorized into key themes, which will be discussed in detail in the following sections.

ID	Database	Title
SS1	Scopus	A model to create a personalized online course based on the student's learning styles
SS2	IEEE	Making DelProctor Proctoring Applications Using OpenCV
SS3	Scopus	Using Learning Analytics to Predict Students Performance in Moodle LMS
SS4	Scopus	Engagement of Students in Data Visualization for the Purpose of E-Learning Improvement
SS5	Scopus	Monitoring Students at the University: Design and Application of a Moodle Plugin
SS6	IEEE	Proposal for the Design and Evaluation of a Dashboard for the Analysis of Learner Behavior and Dropout Prediction in Moodle
SS7	IEEE	Analysis of Teaching and Learning in Moodle with the Help of Visualization with Power BI: Case Study of e-Learning Course of University of Tehran
SS8	IEEE	Activity and Dropout Tracking in Moodle Using UBUMonitor Application
SS9	IEEE	Development of a Data Visualization Assistance System for Online Education Platforms: A Case Study on Moodle

Table 05 - List of Selected Studies

Extraction and Synthesis of Results

The final selection of 9 articles was analyzed by comparing the data with the research questions posed at the beginning of the literature review process, and the results are discussed below.

Overview of Selected Studies

The articles analyzed in this systematic review were published from January 2020 to March 2024. Regarding publication years, there was a uniformity and consistency of publications year by year during the analyzed period. Concerning the types of publication sources, most (5 articles) were published in journals, with the remainder presented at conferences. Another observed point was the classification of the studies as quantitative, with 7 articles falling into this category. Figures 2, 3, and 4 provide an overview of the selected publications.

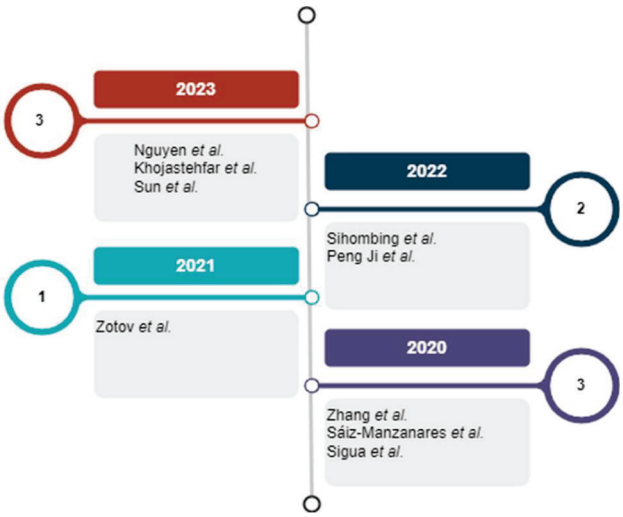

Figure 2 - Timeline

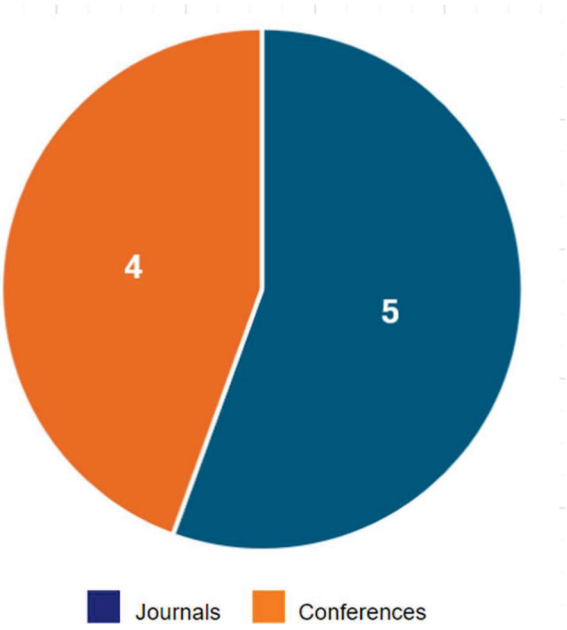

Figure 3 - Publication sources

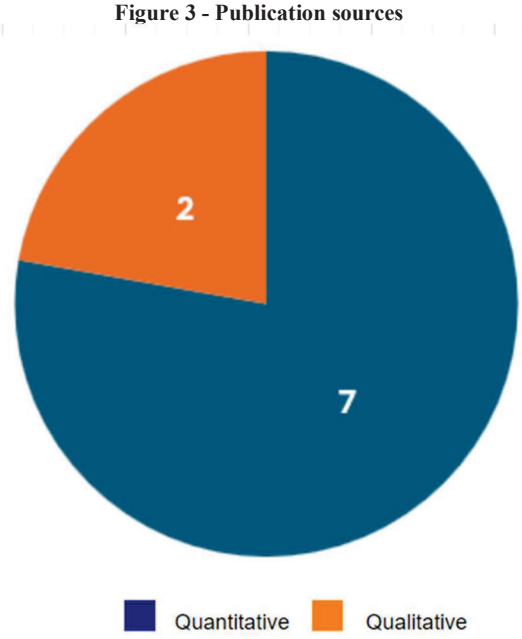

Figure 4 - Types of researches

The analysis indicates a growing interest in this field, with researchers increasingly favoring quantitative methodologies to address emerging issues.

The graphical representations in Figures 2, 3, and 4 provide a visual overview of the selected publications, supporting the conclusion that there is growing interest and an evolution in the methodological approaches adopted by researchers.

Regarding Data Collection Methods

The studies examined utilized the following techniques for data collection:

- **Questionnaires**: Collecting data from interactive activities (views, posts, exercises, forum interactions, etc.) by students on Moodle.
- **SUS Questionnaire (System Usability Scale)**: Evaluating the usability of the system.
- **Moodle Activity Logs**: Recording the time and date of access, IP address, actions completed (viewing, adding, updating, or deleting), logs, click records, time intervals in interactions, and transactions between viewed content.

Regarding the Data Analysis Methods

The studies demonstrate a predominance of statistically validated techniques. Notable methods include:

- **Data Processing**: Pre-processing (cleaning, normalization).
- **Clustering**: K-Means algorithm and Spectral Clustering.
- **Classification**: Correlation analysis, normality analysis using skewness and kurtosis statistics.
- **Data Distributions**: Tested using Analysis of Variance (ANOVA), fixed effect factors (type of degree), parametric tests.
- **Additional Techniques**: Multiple Regression Analysis and decision tree techniques.

Responses to the Research Questions

RQ1: How do the studies present methods for extracting data from Moodle?

Regarding the first research question, the studies identified two primary methods for data extraction, depending on the characteristics of the developed applications. The first, more conventional method is direct database queries. The second method is data extraction through APIs. A criticism related to APIs is the lack of a single API that directly returns all user activity data.

For ready-to-use applications aimed at extraction, the community currently has tools such as the GISMO plugin, which extracts the trail of all activities performed by students and generates graphs (Y. P. Ji et al., 2022). Another tool is MOCLog, which analyzes and presents log data from a

Moodle server (Zotov et al., 2021). The AAT (Academic Analytics Tool) solution aims to provide educators or e-learning course creators with more detailed and meaningful information about students' behavior and the use of learning resources available in Moodle (E. Sigua et al., 2020).

Among the studies addressing this question, the work (SS01) demonstrates that the collected data enabled the identification of students' preferences and their correlation with several learning theories (David Kolb, Honey-Mumford, VARK Method, and Felder-Silverman Learning Styles Model - FLSM), making it possible to adapt presentations.

Among the selected works, some stand out for their results and are outlined below:

Selected Work 02 (SS2) focuses on the creation of a proctoring application called DelProctor, utilizing the OpenCV library. The goal is to develop an efficient and automated solution for monitoring online assessments, ensuring the integrity and authenticity of remote evaluation processes. To achieve this, the work employs computer vision techniques such as video capture, face detection algorithms (Haar cascades), position analysis, focus monitoring, motion detection, and behavior analysis.

Selected Works 03 and 06 (SS03, SS06) demonstrate data prediction techniques used with Moodle. These studies capture data in various forms, including student activity logs, grades, forum interactions, time spent on tasks, and access to study materials. The data are categorized into dependent and independent variables for subsequent processing and report generation, focusing on student behavior on the platform and predicting dropout rates.

RQ2: What are the main dashboard/monitoring tools investigated for use with Moodle?

For the final research question of this systematic review, Question 02 aims to present findings from the selected research that demonstrate how dashboard/monitoring solutions are being utilized based on data extracted from Moodle.

For monitoring purposes, solutions identified include SmartK-lass2, IntelliBoard, Dropout Detective, and UBU Monitor (Y. P. Ji et al., 2022). Regarding visualization dashboards, tools such as Power BI, SQL, Google Charts, and licensed analytical tools are highlighted (Zotov et al., 2021; Sáiz-Manzanares, Marticorena-Sánchez, García-Osorio, 2020; A. Khojastehfar et al., 2023). The selected works (SS4, SS5, SS8, and SS9) stand out in addressing research question RQ2.

Research SS4 presents a visualization panel of activities based on modified state transition networks to analyze and visualize the student's learning path. The student trajectory networks show individual or group students' interactions with the course structure and materials. **Research SS5** features an interactive dashboard that allows teachers to view engagement and performance metrics in real-time. The researchers extracted data mainly from logs (education levels and participation) from various Moodle resources. In addition to dashboards, the work results in automated and periodic reports that highlight at-risk students and provide insights into class behavior. These reports can contribute to implementing an alert system that notifies educators when a student shows signs of risk, enabling proactive interventions.

SS8 and SS9 address the growing need for data-driven tools to enhance online education. Both emphasize the importance of monitoring student engagement and performance and utilizing data visualizations to support informed educational decisions. While the data visualization system has a broader scope in analyzing and presenting various educational metrics, UBU Monitor focuses specifically on tracking activities and preventing school dropout. Together, they demonstrate how technology can transform the management and support of learning in e-learning platforms like Moodle.

Study Limitations

The main limitation of this systematic review is that, even if the defined protocol is followed rigorously, errors may occur in the selection of publications and data extraction, as described in Section 2.4 (Fig. 01), which aims to ensure fairness in the selection process. Another limitation is the search string used, which, despite being defined in terms of research questions and recurring terms in the field, still carries the risk of missing some studies due to the omission of synonyms.

For bibliographic databases and search engines, the main limitation is the indexing time. After an article is published, it takes some time to be indexed, so there is a risk that some recently published research might not yet be included, for example, if it was published in the first quarter of 2024.

Discussion of Results

The systematic review of literature on data collection and visualization, focusing on the Moodle LMS, reveals significant advancements in the analysis and use of data to enhance teaching and learning. The analyzed articles highlight the growing importance of data visualization and activity monitoring tools to promote more effective and personalized education.

The integration of data visualization systems, as demonstrated by Sun et al. (2023), offers a structured and accessible approach for educators to evaluate student performance. The ability to create interactive dashboards and customized visualizations facilitates the identification of trends, student behavior patterns, and areas for improvement in the teaching process. This not only enhances the monitoring of student progress but also empowers educators to adjust their pedagogical strategies more proactively.

On the other hand, the study by Peng Ji et al. (2022) emphasizes the critical importance of monitoring student engagement to prevent school dropout. By using predictive

algorithms and real-time data analysis, UBU Monitor proved to be an effective tool for identifying at-risk students, allowing for early and personalized interventions. This not only potentially reduces dropout rates but also fosters a more inclusive and supportive educational environment.

However, despite significant advancements, the review of the articles identifies ongoing challenges. Implementing these tools often requires substantial resources in terms of technological infrastructure and the training of various actors involved in distance education. Additionally, ethical issues related to student data privacy and the responsible interpretation of collected information are areas of continued concern that were not identified in the studies selected for this review.

Future development of these systems should focus on improving the accuracy of predictive algorithms, integrating additional data sources for more comprehensive analysis, and expanding visualization capabilities to support increasingly complex decisions. Collaboration among researchers, educators, and technology developers is essential to ensure that technological innovations translate into meaningful improvements in online education.

Final Considerations

The reviewed studies provide valuable insights into how data visualization and activity monitoring tools can transform educational management in the context of distance education. As we move towards a future where data analysis becomes a standard practice on platforms like Moodle, we can expect not only improvements in student academic performance but also a more dynamic, adaptable, and student-centered educational environment.

This systematic review highlights the importance of continuing to explore and develop innovative strategies that leverage emerging technologies to promote high-quality education accessible to all.

REFERENCES

1. A. KHOJASTEHFAR et al. Analysis of Teaching and Learning in Moodle with the Help of Visualization with Power BI: Case Study of e-Learning Course of University of Tehran. In: 2023 10TH INTERNATIONAL AND THE 16TH NATIONAL CONFERENCE ON E-LEARNING AND E-TEACHING (ICELET), 28 mar. 2023., journalAbbreviation: 2023 10th International and the 16th National Conference on E-Learning and E-Teaching (ICeLeT). 2023 10th International and the 16th National Conference on E-Learning and E-Teaching (ICeLeT) [...]. [S. l.: s. n.], 28 mar. 2023. p. 1–7.

2. BIOLCHINI, J., MIAN, P.G., NATALI, A.C., TRAVASSOS, G.H., "Systematic Review in Software Engineering: Relevance and Utility", Relatório Técnico ES-679/05, COPPE/UFRJ, Rio de Janeiro, Brasil, 2005.

3. DE OLIVEIRA, P. L. S.; RODRIGUES, R. L.; MACIEL, A. M. A. Revisão Sistemática da Literatura sobre Modelos de Visualizações Utilizados para Analisar Fenômenos Educacionais em Plataformas de LMS. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 16 nov. 2022. Anais do XXXIII Simpósio Brasileiro de Informática na Educação (SBIE 2022) [...]. Brasil: Sociedade Brasileira de Computação - SBC, 16 nov. 2022. p. 1006–1015. Disponível em: https://sol.sbc.org.br/index.php/sbie/article/view/22477. Acesso em: 22 jun. 2024.

4. E. SIGUA *et al.* Proposal for the Design and Evaluation of a Dashboard for the Analysis of Learner Behavior and Dropout Prediction in Moodle. *In*: 2020 XV CONFERENCIA LATINOAMERICANA DE TECNOLOGIAS DE APRENDIZAJE (LACLO), 19 out. 2020., journalAbbreviation: 2020 XV Conferencia Latinoamericana de Tecnologias de Aprendizaje (LACLO). 2020 XV Conferencia Latinoamericana de Tecnologias de Aprendizaje (LACLO) [...]. [*S. l.*: *s. n.*], 19 out. 2020. p. 1–6.

5. KITCHENHAM, Barbara; CHARTERS, Stuart. Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report. 2007. Disponível em: elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf. Acesso em: 05 de abril de 2024.

6. LEMES, T. de C.; DIAS, M. O. de S.; OLIVEIRA, T. de. Análise do uso de dashboard como ferramenta de apoio à tomada de decisão em instituições de ensino: uma Revisão Sistemática da Literatura. Revista Novas Tecnologias na Educação, Porto Alegre, v. 21, n. 1, p. 281–290, 2023. DOI: 10.22456/1679-1916.134356. Disponível em: https://seer.ufrgs.br/index.php/renote/article/view/134356. Acesso em: 10 abr. 2024.

7. MORAN, José Manuel. Os modelos educacionais na aprendizagem on-line. Site pessoal do autor, São Paulo, artigo atualizado em 2002. Disponível em: <

https://moran.eca.usp.br/wp-content/uploads/2013/12/dist.pdf >. Acesso em: 10 de abril de 2024.

8. NAKAGAWA, E. Y. *et al.* Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática. [*S. l.*]: Elsevier Brasil, 2017. Disponível em: https://books.google.com.br/books?id=kCspDwAAQBAJ.

9. PORTO, B.; DIAS, D. M.; BATTESTIN, V. Tendências de Learning Analytics em Moodle: uma Revisão Sistemática. EaD em Foco, [*S. l.*], v. 13, n. 1, p. e2070, 27 nov. 2023.

10. SÁIZ-MANZANARES, M. C.; MARTICORENA-SÁNCHEZ, R.; GARCÍA-OSORIO, C. I. Monitoring students at the university: Design and application of a moodle plugin. Applied Sciences (Switzerland), [*S. l.*], v. 10, n. 10, 2020. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085682183&doi=10.3390%2fapp10103469&partnerID=40&md5=c490c8db1a3f8e842a579be57a35f083.

11. Universidade Aberta do Brasil. Disponível em: https://www.gov.br/capes/pt-br/acesso-a-informacao/acoes-e-programas/educacao-a-distancia/universidade-aberta-do-brasil>. Acesso em: 09 abr. 2024.

12. Y. P. JI *et al.* Activity and Dropout Tracking in Moodle Using UBUMonitor Application. IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, [*S. l.*], v. 17, n. 3, p. 307–317, ago. 2022.

13. ZOTOV, V. *et al.* Engagement of Students in Data Visualization for the Purpose of E-Learning Improvement. International Journal of Emerging Technologies in Learning, [*S. l.*], v. 16, n. 2, p. 46–63, 2021.