Volume 11(3), 82–105. https://doi.org/10.18608/jla.2024.8403

Words of Wisdom: A Journey through the Realm of Natural Language Processing for Learning Analytics—A Systematic Literature Review

Rafael Ferreira Mello[1], Elyda Freitas[2], Luciano Cabral[3], Filipe Dwan Pereira[4], Luiz Rodrigues[5], Mladen Rakovic[6], Jackson Raniel[7], Dragan Gašević[8]

Abstract

Learning analytics (LA) involves the measurement, collection, analysis, and reporting of data about learners and their contexts, aiming to understand and optimize both the learning process and the environments in which it occurs. Among many themes that the LA community considers, natural language processing (NLP) algorithms have been widely adopted to extract information from textual data generated in learning environments (e.g., student essays and short answers, online discussion and chat). NLP can shed light on the learning process and student outcomes in different contexts. Based on the importance of NLP for education, this paper conducted a systematic literature review of the application of NLP to understand how the LA community has been applying the methods from this field. Our methodology includes automatic and manual methods to extract information about authors, relevant papers, and specific data related to educational applications and algorithms used in the field. This review selected 156 papers that reveal essential aspects of the topic; e.g., (i) the majority of the works focused on the analysis of online discussions and essay assessment; (ii) in general, the authors did not apply the developed models in real settings; (iii) recent papers selected have begun to evaluate deep learning models (e.g., BERT) more frequently; and (iv) the datasets used in the experimentation are usually small and contain English text. The results of this study and its practical implications are further discussed.

Notes for Practice

- Natural language processing (NLP) has a critical role in learning analytics (LA) since it can be applied to several educational applications.
- This systematic literature review (SLR) performs a detailed analysis of 156 papers from the perspective of technology and education.
- The main contribution is a set of findings related to the technical aspects of the NLP methods and techniques used in LA and the impact of the NLP-based approaches on the actual learning outcomes.
- This SLR demonstrated a gap between model development and practical use.

Keywords

Natural language processing, text analytics, text mining, computational linguistics, learning analytics, systematic literature review.

Submitted: 01/03/2024 — **Accepted:** 04/10/2024 — **Published:** 30/10/2024

*Corresponding author[1] Email: rflm@cesar.org.br Address: CESAR—Centro de Estudos e Sistemas Avançados do Recife, Recife, PE, Brazil ORCID ID: https://orcid.org/0000-0003-3548-9670

[2] Email: elyda.freitas@upe.br Address: Universidade de Pernambuco, Caruaru, PE, Brazil ORCID ID: https://orcid.org/0000-0001-7439-9040

[3] Email: luciano.cabral@jaboatao.ifpe.edu.br Address: Federal Institute of Pernambuco, Recife, PE, Brazil ORCID ID: https://orcid.org/0000-0002-4235-5753

[4] Email: filipe.dwan@ufrr.br Address: Federal University of Roraima, Boa Vista, Brazil ORCID ID: https://orcid.org/0000-0003-4914-3347

[5] Email: luiz.rodrigues@nees.ufal.br Address: Instituto de Computação, Universidade Federal de Alagoas, Maceió, AL, Brazil ORCID ID: https://orcid.org/0000-0003-0343-3701

[6] Email: mladen.rakovic@monash.edu Address: Faculty of Information Technology, Monash University: Clayton, VIC, AU ORCID ID: https://orcid.org/0000-0002-1413-1103

[7] Email: jackson.florencio@upe.br Address: Universidade de Pernambuco, Caruaru, PE, Brazil ORCID ID: https://orcid.org/0000-0002-4355-7410

[8] Email: dragan.gasevic@monash.edu Address: Faculty of Information Technology, Monash University: Clayton, VIC, AU ORCID ID: https://orcid.org/0000-0001-9265-1908

1. Introduction

Following the growing interest in online education in recent years (Khalil et al., 2021), the number of online learning environments that can support students and instructors in different learning tasks has also increased. Large amounts of textual data—e.g., student essays, blogs, discussion posts, and chat utterances—are created in those learning environments daily (Ferreira-Mello et al., 2019). Whereas textual data may contain rich information about students' characteristics and learning engagement (Azevedo, 2015), they often lack sufficient structure, i.e., textual data is not provided in a form ready for analysis (Balducci & Marinova, 2018). For this reason, textual data is typically more challenging to analyze and extract relevant information from than structured data, which comes in a tabular form (Thomas, 2014). Understanding the textual data could be an opportunity to gain a deeper insight into students' learning processes and progress, which may boost the overall effectiveness of learning technologies and students' and instructors' enthusiasm regarding technology adoption (Castañeda & Selwyn, 2018).

To address the above challenges, the research community has been using different text analytic methods. Commonly, these methods are referred to as text analytics. In particular, text analytics combines methods in natural language processing (NLP), statistical and machine learning, and text visualization to transform and make sense of textual data collected in online learning environments (McNamara et al., 2017).

Despite the promising results of NLP in educational research (Ferreira-Mello et al., 2019), there are instances where studies fail to effectively integrate educational theory with the needs of stakeholders. In this context, learning analytics (LA) has emerged as a vital field. It concentrates on gathering, analyzing, and reporting data about learners and their learning environments (Siemens & Gasevic, 2012). The LA community blends educational theory with data science methodologies to deliver tangible benefits to students, teachers, and institutions (Tan et al., 2016; Aguerrebere et al., 2017).

Recent literature reviews highlight the main trends in applying NLP to solve educational problems. Ferreira-Mello and colleagues (2019) conducted a comprehensive review of this topic and analyzed 343 papers published from 2006 to 2018. The authors provided a broad overview of the major educational tasks (e.g., assessment, student support, question/content generation, and feedback) and main textual products (e.g., online discussion posts, essays, and chat messages) analyzed using text analytical methods. Ahadi and colleagues (2022) conducted another literature review of the topic, specifically focusing on computational methods for text analysis. For instance, the authors found that the five most relevant technical keywords in the field were natural language processing ($N = 140$), sentiment analysis ($N = 131$), machine learning ($N = 122$), text mining ($N = 122$), and deep learning ($N = 64$).

While previous reviews have presented relevant information on the role of NLP in education, there is a lack of systematic reviews that include (i) the specific details (e.g., dataset, methods, and resources) about a large number of NLP tasks for education and (ii) the role of NLP in educational contexts (e.g., educational tasks, textual productions, and impact on student performance). Furthermore, significant advancements in NLP algorithms have happened in recent years that were not covered in earlier literature reviews.

To fill this gap, we conducted a systematic literature review (SLR) on the application of NLP by the LA community. This SLR provides detailed descriptions relative to three major aspects: (i) demographics of the papers (authors, papers per year, countries); (ii) educational goals, techniques, educational resources, and evidence of student performance improvement; and (iii) technical details about the NLP models used (features, databases, algorithms, and performance). This paper aims to systematically consolidate the key findings from prior research to facilitate progress in the field.

2. Method

The SLR developed in this study followed the guidelines proposed by Kitchenham and colleagues (2015). The methodology had five main phases: (i) definition of research questions (RQs), (ii) search and selection of papers, (iii) quality assessment, (iv) data extraction, and (v) results analysis and discussion. Because we are targeting the analysis papers published by the LA community, the review focused on the literature from January 2011 (the first Learning Analytics and Knowledge (LAK) Conference) to December 2023. Figure 1 summarizes each step and the number of articles selected in each phase.

2.1 Research Questions

This SLR aimed to identify and analyze primary studies that apply NLP and LA techniques to address educational challenges. Specifically, our goal was to analyze these publications and extract information about the technical aspects of NLP methods and tools being used in LA research and practice, as well as identify and examine the impact of NLP methods and tools on students and the educational environment.

The first RQ aims to understand the landscape of contributions in NLP-related studies within the LA community, which is essential for several reasons. First, identifying key authors and their geographical locations helps us recognize global trends and regional focuses in this interdisciplinary field. Moreover, analyzing the temporal distribution of these papers provides insights into the evolution of the field. It enables us to trace the development of methodologies, observe shifts in research priorities, and understand the historical context of current practices:

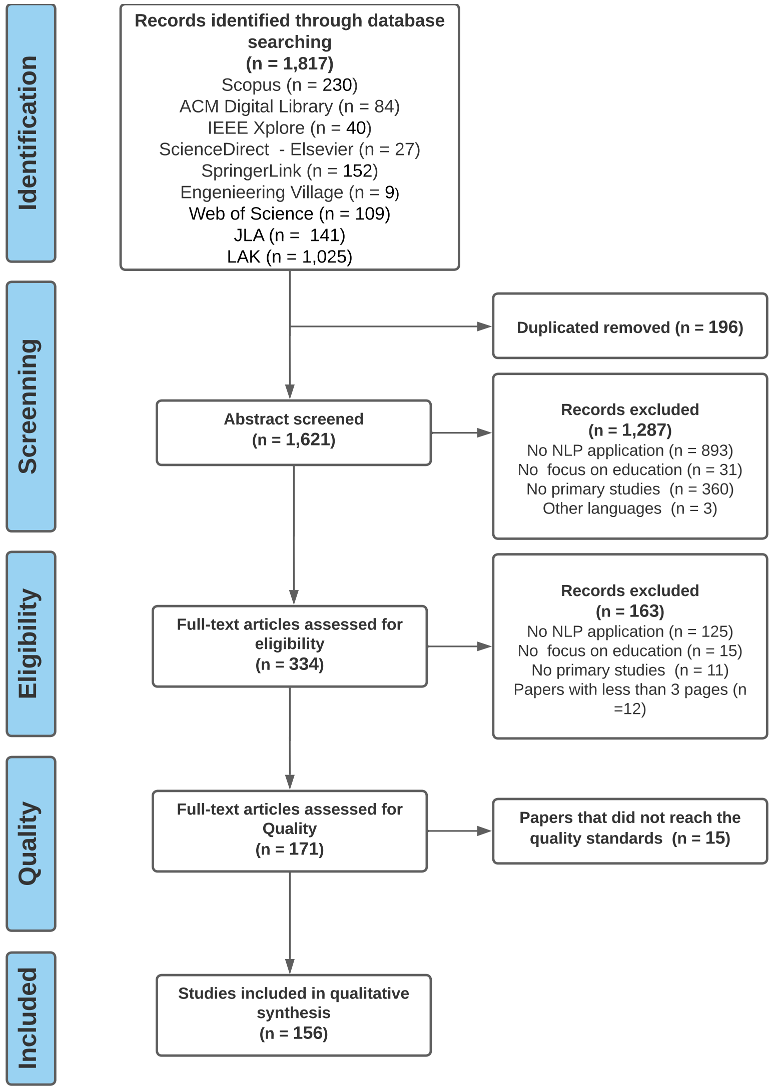


Figure 1. The flow diagram for the SLR detailing each step processed.

RESEARCH QUESTION 1 (RQ1):
Who are the main authors of published NLP-related studies in the LA community? What are their countries? How are the papers distributed over time?

In the second RQ, we aim to map out how NLP techniques are being used to enhance various educational tasks. Understanding these applications will highlight NLP's current capabilities in educational settings and reveal potential areas for further innovation and improvement:

RESEARCH QUESTION 2 (RQ2):
What common educational tasks have been addressed with NLP in LA?

In the context of our study, understanding the specific types of students' textual products analyzed by NLP is crucial. Therefore, the third RQ aims to uncover the diverse range of textual outputs generated by students that are being examined through NLP techniques. By identifying these textual products, we can gain insights into how NLP is being applied to various educational contexts and tasks, such as essay writing, short responses, forum posts, and other forms of written communication:

RESEARCH QUESTION 3 (RQ3):
What main types of students' textual products are being analyzed by NLP?

The fourth RQ evaluates the real-world impact of text analytics on educational outcomes. It offers critical insights into the effectiveness of these tools in enhancing learning, thereby guiding educators and policymakers in technology adoption decisions:

RESEARCH QUESTION 4 (RQ4):
To what extent did the use of text analytics improve student performance, as reported in the reviewed LA studies?

In this review, we also gathered information on how to advance NLP within the educational context. Initially, we emphasize the need to detail the datasets used, including their size, target stakeholder, and educational context. This RQ seeks to uncover the most prevalent units of analysis and typical sample sizes of the dataset used in the field. Additionally, it aims to identify the target population of these studies (e.g., students, teachers, or institutional leaders). As such, our fifth RQ is as follows:

RESEARCH QUESTION 5 (RQ5):
What are the primary characteristics of datasets used in this field?

Then, we also evaluate the main machine learning algorithms and features. The sixth RQ aims to elucidate the methodological application to the LA community by identifying the specific NLP algorithms and feature sets employed. This understanding is crucial for several reasons: first, it allows us to assess the effectiveness of different NLP techniques in educational contexts, guiding future research and application choices. Second, it identifies potential gaps in current methodological approaches, suggesting areas for innovation and development. Finally, understanding the prevalent NLP algorithms and features in LA research can help standardize methodologies, which is essential for comparative studies and advancing the field:

RESEARCH QUESTION 6 (RQ6):
Which NLP algorithms and feature sets have been used to support LA research?

Finally, the last RQ presented a comparative performance of the NLP algorithms in terms of Cohen's kappa and accuracy. It seeks to quantify the effectiveness of NLP algorithms within the context of LA, offering a clear perspective on effectiveness. Such insights are invaluable for practitioners and researchers considering adopting or developing NLP techniques in educational settings. Furthermore, understanding these outcomes facilitates a comparative analysis of different NLP approaches and can set the quality threshold for future studies:

RESEARCH QUESTION 7 (RQ7):
What key outcomes are achieved by the NLP algorithms in the LA studies identified?

2.2 Search Strategies

After defining the RQs, the second step in the systematic review was keyword search. We queried seven academic databases: ACM Digital Library, IEEE Xplore, ScienceDirect, Web of Science, SpringerLink, Engineering Village, and Scopus. Furthermore, we manually evaluated all the papers published in the *Journal of Learning Analytics* (JLA) and LAK. The selection of these databases was based on the literature (Sousa et al., 2021) and assured the inclusion of the conference and journal maintained by the Society for Learning Analytics Research—SoLAR[1]—as the most prominent specialized publication venues for research in LA. The search in all the databases was performed on January 11, 2024.

The query *"learning analytics" AND ("natural language processing" OR "text mining" OR "text analytics" OR "computational linguistics" OR "LLM" OR "large language model")* was applied to each academic database cited above to search. The keywords were applied for the title, abstract, and keywords of the papers in the search engines. The initial search retrieved 1,817 papers.

2.3 Selection Process

To select the papers of interest for this research, we defined the following inclusion criteria: (a) the papers report on NLP applications for education and (b) the papers are either full or short (i.e., posters and papers from workshops and demonstrations were not included). Moreover, we also adopted the following exclusion criteria: (a) secondary studies (i.e., we did not include literature reviews (G. Chen et al., 2020) or theoretical studies (Ferguson & Shum, 2012)) and (b) papers working with whole textual products (e.g., online discussion thread) but not exploring the information within a textual product (e.g., using posts reply) (McAuley et al., 2012).

Initially, two experts in the field read the title, abstract, and keywords of each retrieved paper and decided whether a paper should be included or not. The experts achieved a level of agreement of 89.61% (Cohen's $\kappa = 0.87$) with a total of only 176 disagreements, which were included in the next phase. In the second phase, the experts read the full papers to analyze the

[1] https://www.solaresearch.org

pertinence of this SLR; in this stage, their level of agreement reached 90.71% (Cohen's $\kappa = 0.89$) with a total of only 30 disagreements, which were resolved through discussion. The search resulted in 163 papers selected for further analysis, which included the quality assessment.

2.4 Quality Assessment

After the paper selection, we performed a quality analysis, answering six questions concerning the papers' methodology. The questions were the same as the ones formulated by Sousa and colleagues (2021), except the seventh question, which was specifically related to the application of LA in schools. The questions are presented in Table 1.

Table 1. Questions used to evaluate the quality of the selected studies. Adapted from Sousa and colleagues (2021).

ID	Type	Question
Q1	Bias	Did the study present a research project and not an expert opinion?
Q2	Bias	Did the study fully describe the context analyzed?
Q3	Internal Validity	Were the objectives of the study clearly defined?
Q4	Internal Validity	Was the proposed methodology adequate to achieve the research objectives?
Q5	Internal Validity	Were the data collection methods used and described correctly?
Q6	External Validity	Have the research results been properly validated?

Regarding the process, two researchers read the papers previously selected and answered the questions using "yes" (1 point) or "no" (0 points). Only those papers that achieved 12 points (6 for each researcher) proceeded to the next stage of this SLR. Disagreements were solved through discussion, and only three divergences remained, which were resolved by a third researcher. In the end, from a total of 171 papers, 15 were removed in this step.

2.5 Extraction of Relevant Fields

The next step of this research was data extraction. Three experts manually extracted the following attributes from the papers[2]: (i) bibliographic information: data about the paper publication, such as authors, title, year, and type (full or short paper); (ii) educational aspects: the paper's goal, types of textual products used for analysis, and if the proposed approach was applied in practice or not; (iii) dataset: information about the data used in the papers, including language, the unit of analysis, size, and type of course, among others; (iv) feature engineering: the tools/techniques for feature engineering (e.g., Coh-Metrix, LIWC, and TF-IDF) applied in the studies reported in the papers included in the review; and (v) NLP approaches: the machine learning algorithms used in the papers and the algorithms that had the best performance.

During the process of extracting the above information, the experts initially processed 5% (randomly selected) of the included papers to assure the reliability of the data extraction. A third expert assessed the results and identified the need to better align the concepts of the information to be extracted from the papers. Then, the initial three experts did a new review round with another 5% (randomly selected) of the papers. At this point, the researchers agreed with all the information extracted in the papers of this round. Based on this, the remainder of the papers were evenly divided between the three experts, who each extracted the required data from the papers they were assigned.

3. Results

3.1 RQ1: Quantitative Analysis

The first RQ focused on the descriptive analysis of the selected papers. This review retrieved 156 relevant articles, categorized into 112 full papers and 44 short papers, on the related topic. Figure 2 illustrates the distribution of these papers over the years 2011 to 2023. The data shows a steady increase in the number of publications in NLP since 2015, peaking in 2023, with the exception of lower publication numbers in 2013 and 2018.

The retrieved papers were written by a total of 444 authors, with a mean of 3.98 (3.90) authors per paper, with only five papers with a single author. Figure 3a shows the authors (20) with at least four papers published. It includes senior authors in the field of educational technology and NLP, such as Dragan Gašević, Danielle McNamara, and Simon Buckingham Shum; mid-career researchers, such as Laura Allen, Rafael Ferreira Mello, and Nia Dowell; and early-career researchers, such as Vitor Rolim and Ágnes Sándor. Furthermore, Figure 3b shows the top 10 countries, considering the first author of each paper, which include the US (36.1%), Australia (9.0%), the UK (7.7%), Canada (6.5%), and Brazil (6.5%), with more than 5% of the papers. The figure grouped countries with less than 2% of papers in the OTHER category.

[2] For more details about the attributes extracted, see the link https://shorturl.at/bMdv1.

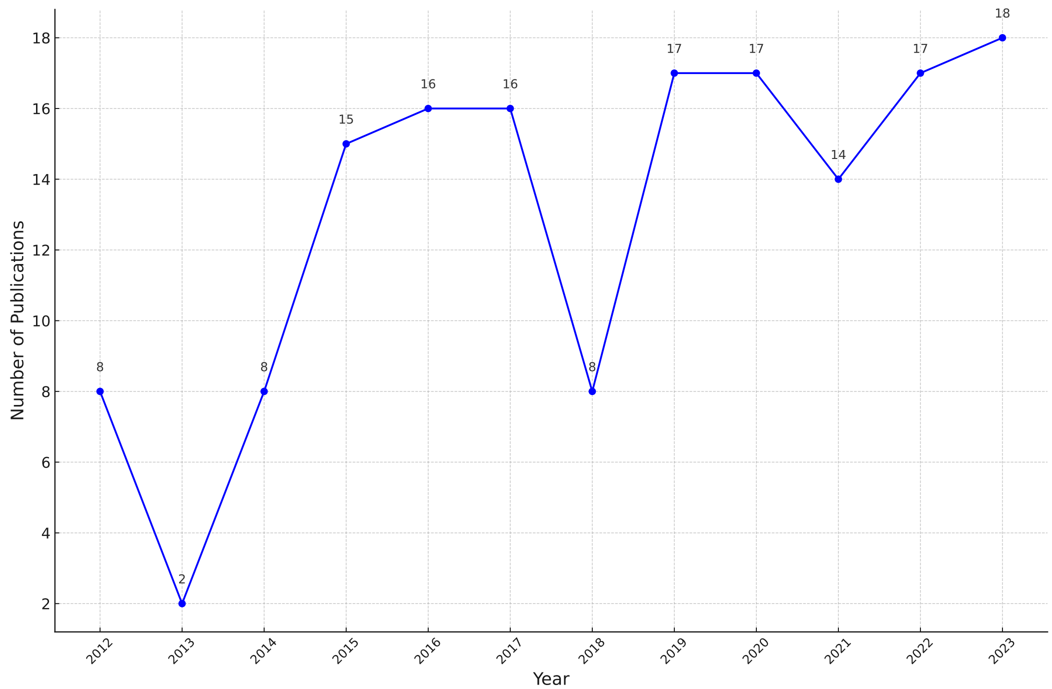


Figure 2. Number of selected papers by year.

3.2 RQ2: Educational Tasks

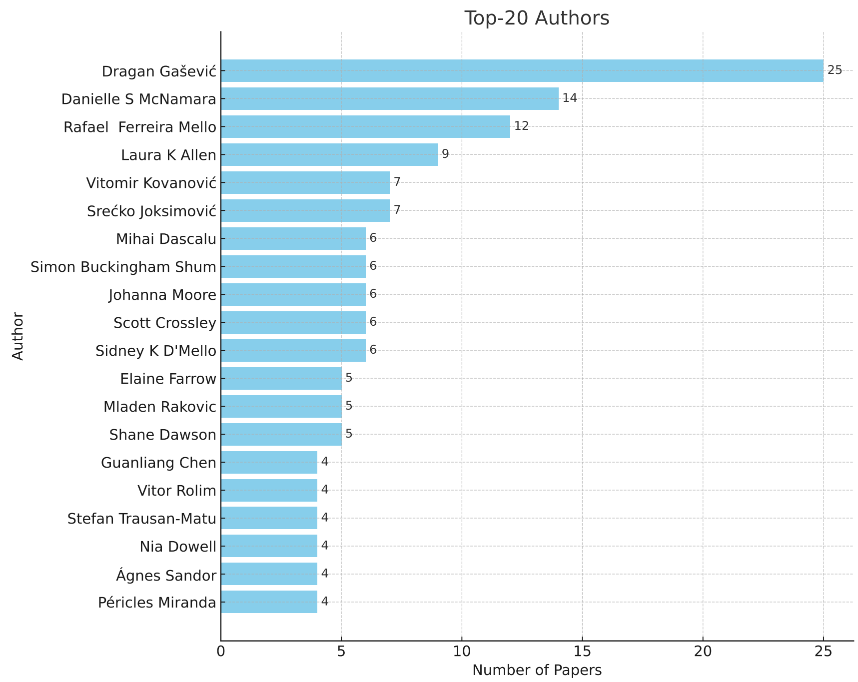
The results of RQ2 present details of educational tasks and textual products that have been investigated/supported by text analytics in LA. Collaborative learning was the most prominent task across the studies reviewed ($N = 49$; Figure 4). Next, 43 studies focused on assessment, 35 studies on feedback, seven on recommendation systems, and five on student modelling and self-regulation. Educational tasks in the remaining studies included reading comprehension, outcome prediction, curricula analysis, and privacy.

Among the studies that reported collaborative learning tasks, the main task was to predict the behaviour of a student or a group of students based on the text data collected from online discussions. To illustrate, many papers proposed content analysis methods to automatically identify social and cognitive presence in online discussions (Kovanović et al., 2016; M. Ferreira et al., 2020; Y. Hu et al., 2020; G. Barbosa et al., 2020). In the same direction, Iqbal and colleagues (2022) used latent Dirichlet allocation (LDA) and block hidden Markov models (HMMs) to identify speech acts from textual data about online discussion messages and then investigated the relationships between cognitive presence and speech acts using a network-based approach and statistical tests. In addition to analyzing online discussions, collaborative learning was also investigated in different assessment settings, such as collaborative writing (Southavilay et al., 2013) and the evaluation of open-ended questions (Crossley et al., 2017; Cross et al., 2017).

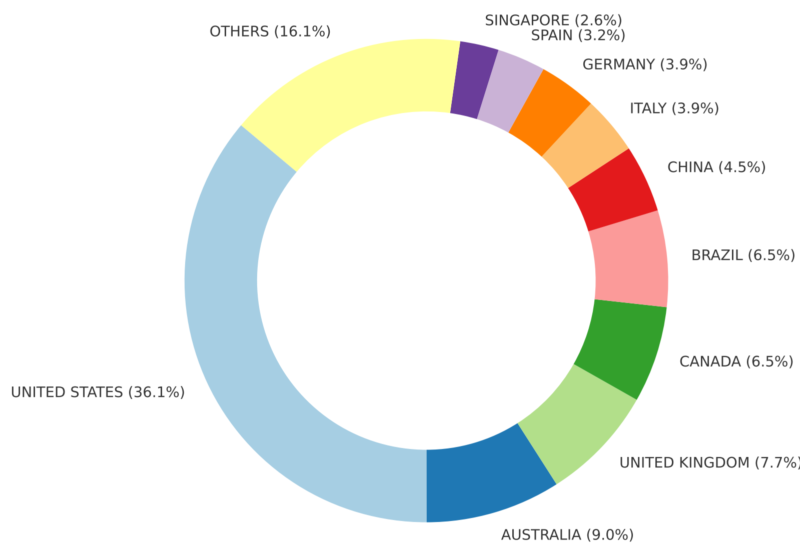
Regarding the use of NLP to support educational feedback, we identified two major areas of focus: the content analysis of feedback messages and the development of algorithms for generating feedback. For instance, Nicoll and colleagues (2022) outlined a method to automatically extract several indicators of instructor-provided feedback and determine how the extracted characteristics correlated with changes in student grades. Similarly, Cavalcanti and colleagues (2020) assessed the quality of instructors' feedback using linguistic features and supervised machine learning. On the other hand, Rüdian and colleagues (2022) explored the usefulness of the automatic generation of feedback in the context of language learning, and Öncel and colleagues (2021) explored ways to personalize feedback generation.

In terms of assessment, the majority of the studies focused on automatic grading of the student's written products on different tasks, such as answers to open-ended questions in math (Erickson et al., 2020), second language acquisition (Zylich & Lan, 2021), and essays (Ferreira Mello et al., 2022). Other examples of studies on assessment included tracking students' knowledge change (Ruan et al., 2021) and expertise on specific subjects (Benedetto et al., 2020) by analyzing how students answered open-ended questions.

The papers about recommendation systems and student modelling provide resources deemed relevant to students (Niemann et al., 2012) or educational stakeholders (Fiallos & Ochoa, 2019) and extracted relevant information from students' textual products (e.g., online discussion messages (Teplovs et al., 2011) or tutorial dialogue (Zylich & Lan, 2021)) to support instructors in providing feedback. Finally, the other category included papers related to the evaluation of reading comprehension skills (Allen et al., 2015), predicting motivation and utility in student responses to open-ended questions to prevent dropout (Robinson et al., 2016), and the analysis of computer science curricula (Sekiya et al., 2015).



(a) The most active authors in NLP for LA.



(b) The most active countries in NLP for LA.

Figure 3. Number of publications per author and country.

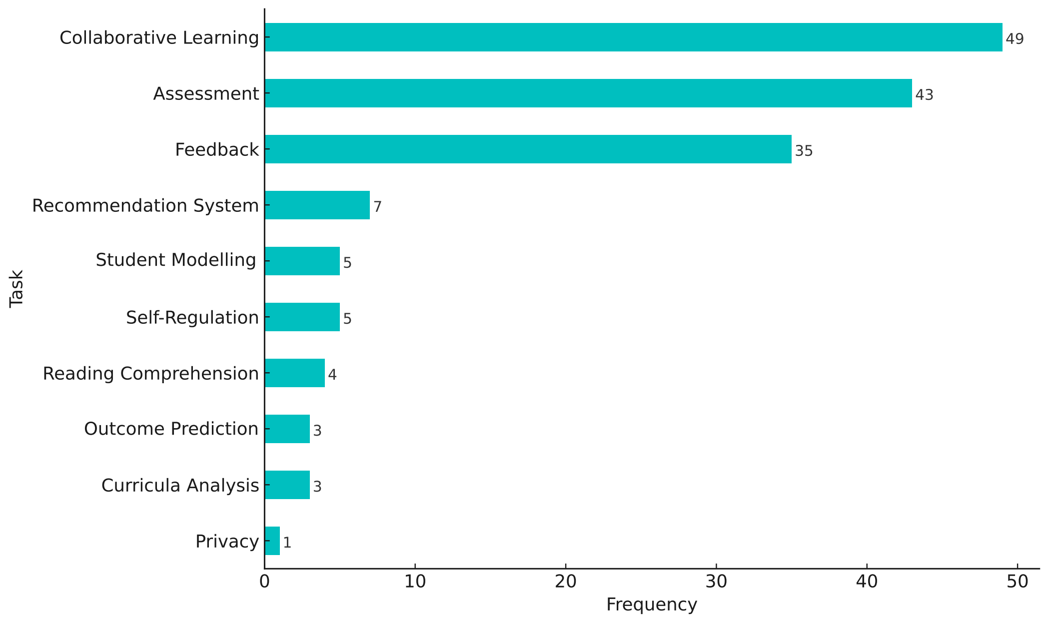


Figure 4. The educational tasks when using NLP in studies in LA.

3.3 RQ3: Textual Resources

This RQ also attempted to understand what types of textual products have been analyzed by the field. Figure 5 illustrates the range of textual resources used in the field. The predominant resource is online discussions (55), followed by essays (26). Additionally, questions (10), short answers (10), feedback messages (9), and reflective statements (9) are also extensively adopted. Other utilized resources include speech transcription, course syllabi, tutorial dialogue, and learning objects.

Understanding the interactions between students and instructors in online discussions has been a key topic in the field of LA. For instance, our results showed that collaborative learning has been the most frequent educational task across the studies reviewed. Moreover, many papers focused on the automatic extraction of relevant indicators of cognitive and social processes from the students' messages in educational forums (Kovanović et al., 2016; M. Ferreira et al., 2020; Liu et al., 2023; G. Barbosa et al., 2020). In addition, we identified studies that reported on the detection of key topics in online discussions (Hsiao & Awasthi, 2015; Lee & Tan, 2017), knowledge creation and student self-reflection (B. Chen & Resendes, 2014; Jung & Friend Wise, 2020), community building (Teplovs et al., 2011; Dascalu et al., 2015), and the evaluation of online discussion participation to predict student performance (Paredes & Chung, 2012; Crossley et al., 2016; Shibani et al., 2016).

In analyzing student essays, we identified two main lines of work: (i) automated essay scoring based on linguistic properties of text and (ii) provision of analytics-based feedback on student writing. Within the first line of studies, we identified the papers that reported on extracting rhetorical moves (Ferreira Mello et al., 2022), writing style (Snow et al., 2015), and originality (Lárusson & White, 2012; Öncel et al., 2021). In the second line, the main focus was providing visualization for actionable feedback (Gibson et al., 2017; Ullmann, 2017; Buckingham Shum et al., 2016). Furthermore, we also identified studies that evaluated student essays and used this information to predict student success (Stone et al., 2019) and to support students during the writing process (Whitelock et al., 2015; Allen et al., 2016).

Analysis of short answers and questions formed another group of studies. The papers about short answers attempted to evaluate students' responses to different disciplines: math (Buckingham Shum et al., 2016; Erickson et al., 2020), sciences (Leeman-Munk et al., 2014), and general-purpose questions (Qiao & Hu, 2019). Regarding the papers related to the analysis of the questions, in general, they are related to the evaluation of question quality (Benedetto et al., 2020; Jayakodi et al., 2016) and management of repositories of question items (Albano et al., 2022, 2023).

The application from the studies selected also focused on identifying reflective elements in student discourse, as highlighted in studies by Jung and Friend Wise (2020), Kovanović and colleagues (2018), and Li and colleagues (2023). These studies emphasize the importance of detecting such elements to enhance student feedback (Gibson et al., 2017). Additionally, the quality of feedback messages has become a focal point in educational discourse analysis. Notable contributions in this area include works by Cavalcanti and colleagues (2020) and Nicoll and colleagues (2022), who have delved into the intricacies of assessing and improving the effectiveness of feedback.

Another group of papers is focused on analyzing transcripts from audio or video data. In this direction, Carnell and colleagues (2019) and Sherin (2012) used NLP to assess communication skills based on one-on-one interviews. It is also

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License (CC BY 4.0)

important to highlight that some papers related to the examination of speech also relied on acoustic and verbal features to enhance the performance of machine learning algorithms for text analysis (Hunkins et al., 2022; Donnelly et al., 2017). Finally, other researchers analyzed teamwork (Swiecki & Shaffer, 2020) and tutorial dialogue (Zylich & Lan, 2021), evaluated program competencies or courses based on the course syllabi (Sekiya et al., 2015; Chang et al., 2022), and studied the adoption of ChatGPT to provide chatbots to support students in different contexts (Qureshi, 2023; Rajala et al., 2023).

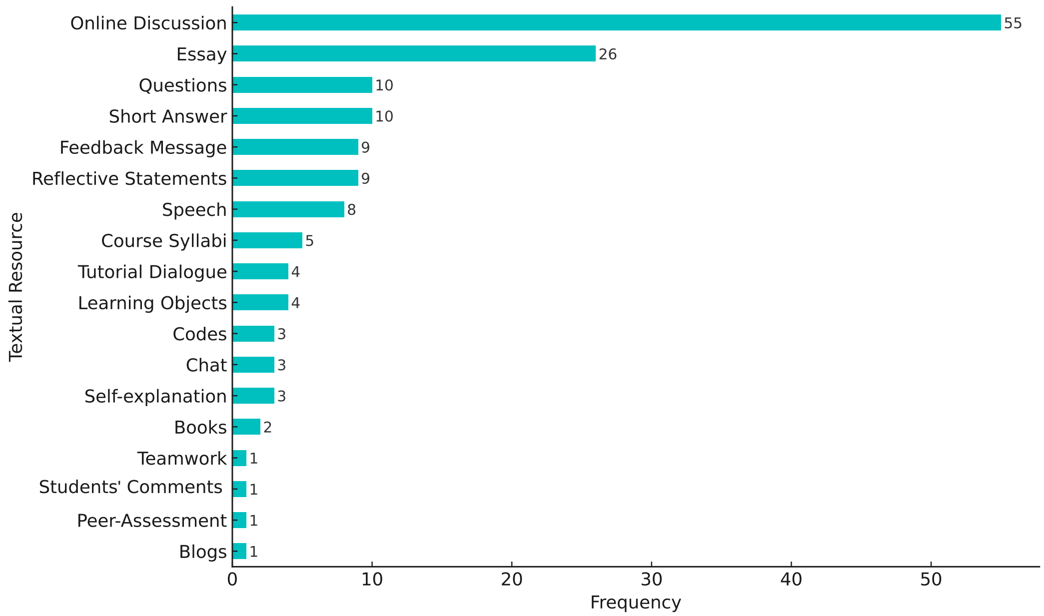


Figure 5. The textual resources used in the field.

3.4 RQ4: Evidence about Support for Learners

This section outlines the application of text analytics in practical settings in the studies reported in the selected papers. In this study, we used the term *evidence* as the empirical scientific indication that text analytics improved learning outcomes or supported learning and/or teaching (Ferguson & Clow, 2017), using the same categories proposed by Sousa and colleagues (2021). Table 2 summarizes the results. It shows that most of the papers included in the review (139—89.10%) did not apply their proposed approach in practice. That is, they mostly focused on the analysis of textual data, but they did not use the results of their analysis in real-world applications. Moreover, six papers (3.84%) assessed the practical application of a text analytic algorithm, but they did not report any evidence to support the impact on learning outcomes and processes.

Table 2. Practical application of NLP in the studies reported for LA.

Evidence	Number of articles (%)
No evidence	6 (3.84%)
Positive with empirical evaluation	5 (3.22%)
Positive without empirical evaluation	6 (3.84%)
Papers evaluated only with previous data	139 (89.10%)
Total	**156 (100.00%)**

The majority of papers that reported a positive impact on learning outcomes or processes presented the practical use of tools based on text analytics to support students in creating written products (Shibani et al., 2019; Whitelock et al., 2015; Southavilay et al., 2013; Gibson et al., 2017; Snow et al., 2015; Lárusson & White, 2012; Knight et al., 2018). For instance, Shibani and colleagues (2019) and Gibson and colleagues (2017) used the tools and visualizations of writing analytics to provide feedback for students. Text analytics have also been applied in practice, demonstrating a positive impact on students' learning outcomes in tasks related to collaborative writing and reading comprehension (Southavilay et al., 2013).

Another aspect evaluated in practice was the potential of using NLP to develop chatbots (Schlippe & Sawatzki, 2022; Rajala et al., 2023; Qureshi, 2023). In the study proposed by Schlippe and Sawatzki (2022), the authors proposed the development of a chatbot to support students' preparation for exams, including gamification aspects. Moreover, recent works (Rajala et al., 2023;

Qureshi, 2023) adopted ChatGPT as a chatbot system to support first-year students in computer science courses (Rajala et al., 2023) and as a programming language tutor (Qureshi, 2023).

Finally, NLP was also used to support interactions in online discussion and chat (Shusterman et al., 2021). In this direction, Shusterman and colleagues (2021) used text analytics to identify relevant messages from previous online discussions (of the same course) that could trigger discussion in a subsequent run of the course.

3.5 RQ5: Description of the Datasets Used

As students and instructors interact with different types of textual resources across learning tasks, different approaches are often needed to analyze those artifacts effectively and purposefully. It is often not sufficient to use NLP algorithms, but it is equally important to define the appropriate unit of analysis (e.g., discussion post vs. paragraph vs. sentence). Therefore, the dataset is a critical aspect for the success of the textual analysis, mainly when using NLP and machine learning techniques. In this section, we aim to describe the datasets used by the selected papers divided among educational textual resources (see Section 3.3) and unit of analysis. Table 3 summarizes the statistics related to the corpora used by the selected papers.

The first part of Table 3 shows the statistics for the most used textual resources extracted from the selected papers. It is possible to see that the mean number of records is usually very high. It is important to mention that the papers used different records as their units of analysis (e.g., full text and words). For instance, the papers for second language learning through tutorial dialogue (Ruan et al., 2021; Zylich & Lan, 2021) using the Duolingo dataset (Settles & Meeder, 2016) had an extensive number of records (1,280,000 sentences). In this case, the median and min-max values could be more insightful to better understand the datasets used.

Online discussions, the most popular textual product used in the selected papers, had a median of 1,747 and minimum and maximum values of 60 and 175,036, respectively. Regarding essays, the second most used textual product had fewer instances per dataset, including one paper evaluating only 25 texts. Regarding the median, speech and feedback messages have the most records. In the first case, the number of records was enormous because the papers considered each speech unit as a single data point (Molenaar & Chiu, 2015; Sherin, 2012). In the second case, the number of records was enormous because many papers considered the analysis of peer review feedback, which tends to have large datasets (Dood et al., 2022; Kong et al., 2023).

Table 3. Descriptive statistics of the datasets used for analysis in the selected papers. It includes mean, standard deviation, median, minimum, and maximum number of instances divided among textual products and units of analysis.

		Number of Unit of Analysis			Number of Participants		
	Number of Papers	Mean (SD)	Median	Min–Max	Mean (SD)	Median	Min–Max
Textual products							
Online Discussion	55	10,369.89 (26,845.81)	1,747	60–175,036	616.45 (1260.64)	275	13–7,814
Essays	26	12,059.56 (50,374.97)	649	25–253,306	511.22 (955.12)	98	22–3,998
Question	10	13,556.30 (35,034.40)	477	53–112,526	17,364.66 (28,280.57)	2,043	51–50,000
Short Answer	10	31,112.57 (55,826.52)	824	207–150,447	10,063 (17,278.24)	331	24–41,946
Feedback Message	9	51,822.87 (99,068.97)	4,246	350–288,312	1,961.20 (3,137.57)	456	2–7,482
Reflective Statements	9	3,682.12 (7,177.04)	1,330	120–21,096	107 (127.19)	30	17–369
Speech	8	9,835.28 (11,597.80)	10,000	36–31,533	121.12 (150.73)	61	11–423
Others	28	1,770,877 (4,156,979)	9,499	180–1,280,000	31,683 (62,480)	154	10–164,196
Unit of Analysis							
Full text	129	12,566.69 (37,307.17)	1,330	15–288,312	3,901.60 (18,434.33)	155	2–16,4196
Sentence	19	693,361.63 (2,932,305.84)	6,972	71–12,800,000	9,086 (31,891.07)	200	11–115,222
Words	4	812,978.33 (1,384,837.57)	24,530	2,405–2,412,000	96 (96.16)	96	28–164

The bottom part of Table 3 shows the number of records in the datasets according to the unit of analysis of each paper. Most papers focused on analyzing all of the textual products ($N = 129$), with significant variability between them (STD $= 37,307.17$). In this case, the median is 1,330, using up to 288,312 textual productions. There are 19 papers adopting sentences as the units of analysis. The median (6,972) for this group is more than five times that of the full text group. Furthermore, the papers using words as units of analysis had the biggest collection of instances, with a median of 24,530 and a maximum of 2,412,000 words analyzed. In the remaining studies, authors used paragraphs (Allen et al., 2019), stanzas of speech (Molenaar & Chiu, 2015), and topic segments (Knight & Littleton, 2015) as units of analysis.

Two important additional aspects about the datasets are as follows: (i) 115 out of the 156 papers evaluated only English texts. The other papers considered Arabic, Chinese, Dutch, Portuguese, French, German, Italian, Romanian, Greek, Mandarin, and Spanish. (ii) Only five papers provided the dataset as open access, limiting the reproducibility of the results.

3.6 RQ6: NLP Algorithms and Feature Sets

Our sixth RQ aimed to identify the most prominent NLP algorithms and feature sets used by the LA community to support text analysis. Figure 6 shows the 10 most utilized algorithms over time. This analysis is more relevant than the simple volumetry analysis because some models (e.g., BERT) were developed recently. The results reveal a general trend toward using conventional machine learning algorithms, especially until 2017, with the widespread use of deep learning algorithms (i.e., the algorithms based on deep neural networks). However, LSTM and BERT have been widely used since 2021. Moreover, in the studies included in the review, the authors widely adopted the white-box models (e.g., Decision Tree, Linear and Logistic Regression, and Naive Bayes), i.e., the models that allow for more straightforward interpretability of the prediction results. Regarding specific machine learning algorithms, the Random Forest classifier was the most commonly used ($N = 30$), whereas Logistic Regression and SVM were second in this list ($N = 16$). The current analysis shows that Random Forest and Logistic Regression are the more stable algorithms that reached higher occurrences. Furthermore, BERT, LSTM, and XGBoost demonstrate an increase in adoption.

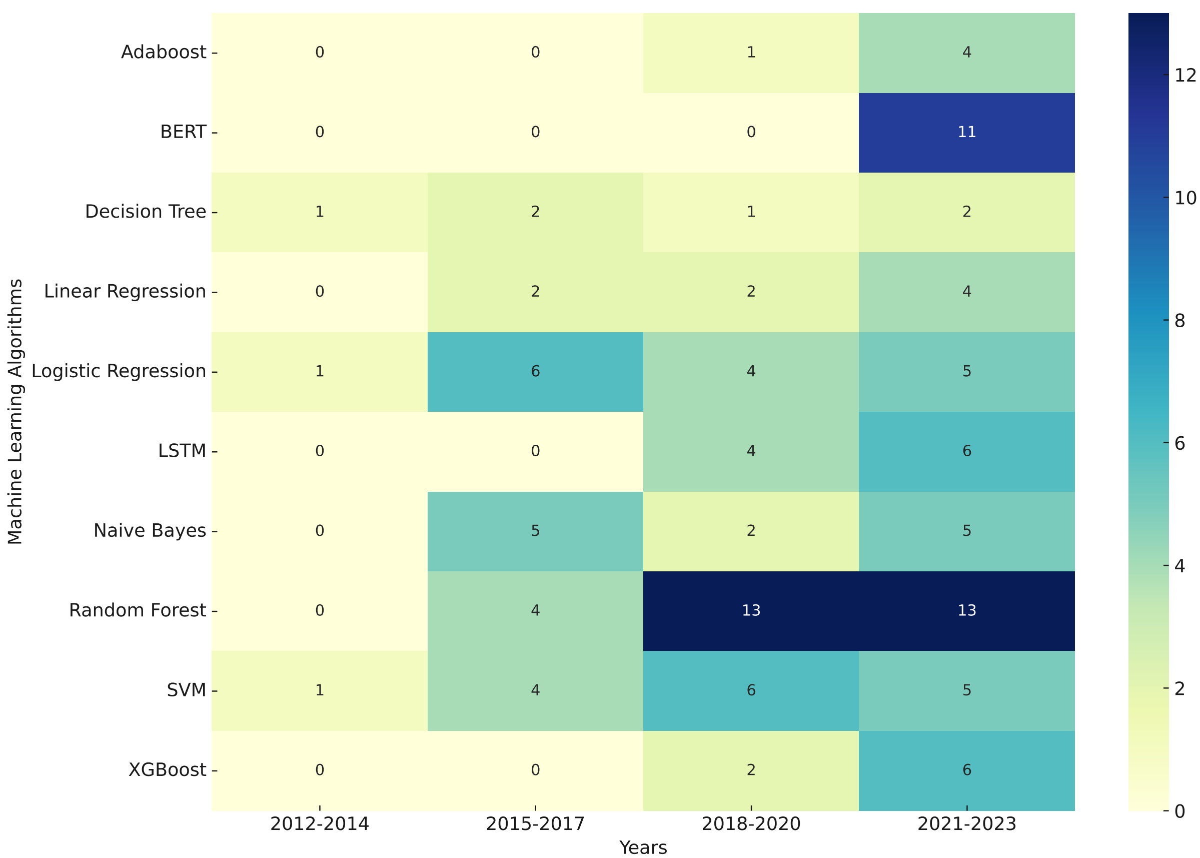


Figure 6. The most commonly used machine learning algorithms in the studies that used NLP for LA over the years.

The main techniques found were supervised machine learning algorithms ($N = 90$), meaning that textual data was automatically categorized in a set of pre-defined categories (classification) or a scalar number (regression). In addition to

ISSN 1929-7750 (online).

the final accuracy of a model, many papers also reported on the most important features in different classification algorithms (Kovanović et al., 2016; M. Ferreira et al., 2020; Y. Hu et al., 2020; G. Barbosa et al., 2020; Ferreira Mello et al., 2022). The majority of the deep learning models were applied to multi-modal data, e.g., speech (Pugh et al., 2022) and game environments (Park et al., 2021), with a broader adoption (after 2019) in the analysis of online discussion and open-ended questions (Shusterman et al., 2021; Dood et al., 2022; Ruan et al., 2021; Erickson et al., 2020).

The use of unsupervised machine learning algorithms ($N = 28$) focused on providing an exploratory analysis of textual data (Sherin, 2012), clustering students based on their messages in online discussion (M. A. D. Ferreira et al., 2022), creating students' models (B. Chen & Resendes, 2014), extracting topics from texts (Vytasek et al., 2019), and using similarity methods (Albano et al., 2023). It is important to mention that among the selected studies, 35 did not use machine learning methods. In this case, the papers extract indicators based on linguistic features (Paredes & Chung, 2012; Whitelock et al., 2015), perform text similarity measures to cluster educational resources (Niemann et al., 2012), and adopt different methods as epistemic network analysis (Fougt et al., 2018), among others.

Figure 7 reveals the most commonly used features in the studies reported in the selected papers over time. Of all features, TF-IDF ($N = 40$), the most commonly used in text analytics (Chowdhary, 2020) in general, including education (Ferreira-Mello et al., 2019), was the most popular. It keeps steady over time as it is largely used as a baseline in NLP studies. However, it is crucial to emphasize that, over the past three years (2021–2023), word embeddings, particularly BERT (Del Gobbo et al., 2023; Dood et al., 2022; Liu et al., 2023), have been the predominant approach, used in 28 studies. This frequency is more than using TF-IDF and structural features combined during the same period.

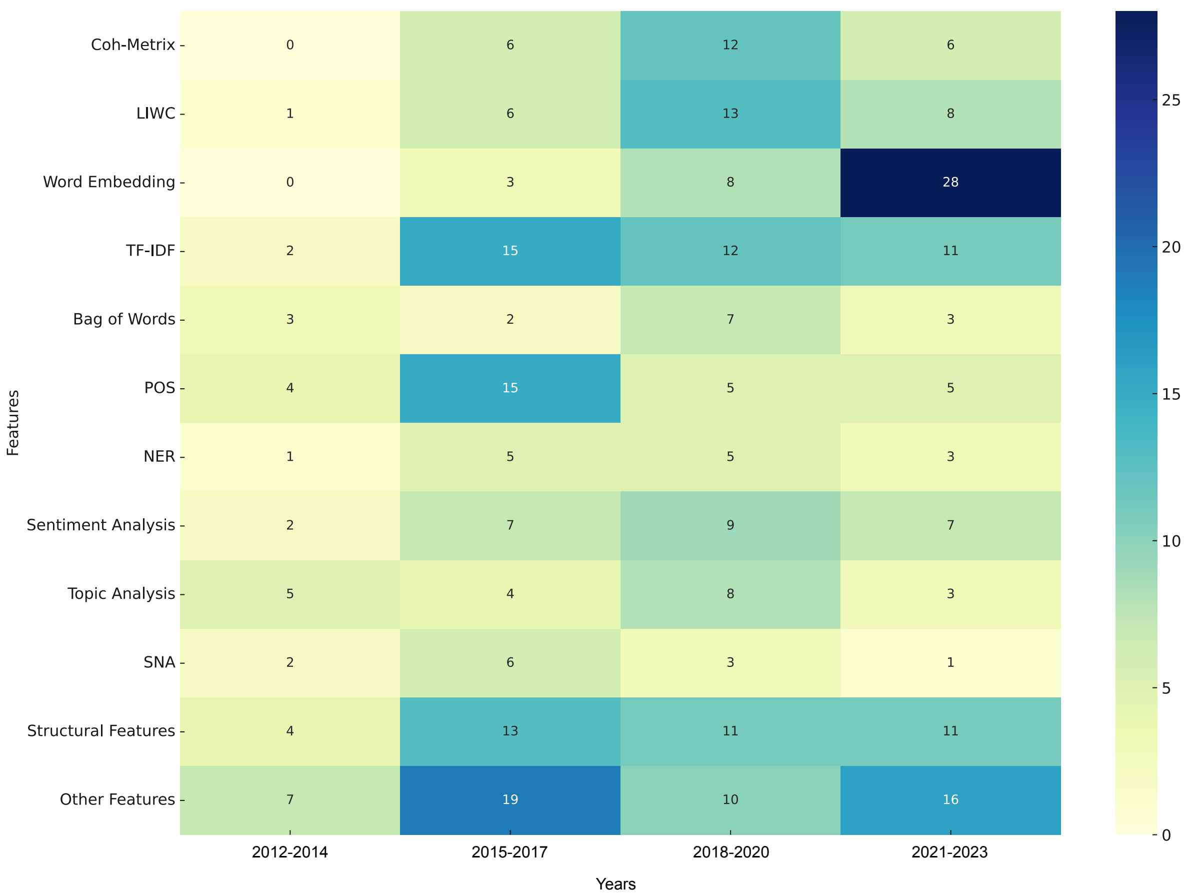


Figure 7. The most commonly used features in studies selected.

Following TF-IDF and word embedding, structural features (e.g., discussion context features and features based on syntactic analysis) (Kovanović et al., 2016), LIWC (Tausczik & Pennebaker, 2010), Coh-Metrix (McNamara et al., 2014), part-of-speech tags (POS) (Chowdhary, 2020), sentiment analysis (Nicoll et al., 2022; Park et al., 2021), latent semantic analysis (LSA) (Chowdhary, 2020; Raković et al., 2021; Rakovic et al., 2022; Raković et al., 2023), name entity recognition (NER) (Chowdhary, 2020), traditional bag of words (BoW) (Chowdhary, 2020), and social network analysis (SNA) (Serrat, 2017) have been widely used to extract indicators in the studies that included NLP in LA. The other category encompassed features that appeared less than five times, including entropy measures (Snow et al., 2015), ordering features (Ferreira Mello et al., 2022), co-occurrence of words (Niemann et al., 2012), and other linguistic indexes (e.g., Automated Analyses of Lexical Sophistication (TAALES), Automated Analysis of Cohesion (TAACO) and Cognition Engine (SEANCE)) (Paredes & Chung, 2012; Hunkins et al., 2022; Donnelly et al., 2017; Allen et al., 2019).

3.7 RQ7: Better Outcomes

The final RQ we propose seeks to elucidate the impact of NLP algorithms within the LA domain. This question encompasses two primary analytical dimensions: (i) identification and evaluation of the most effective NLP algorithms yearly and (ii) a comprehensive synthesis of key results from selected papers focusing on Cohen's kappa and accuracy.

Figure 8 reveals the algorithms that reached the best performance in previous studies over the years. While this analysis connects to RQ6, it offers distinct insights by concentrating on the leading algorithm from each study. The figure indicates that Random Forest was the predominant algorithm in the initial years, particularly peaking in popularity in 2020. In contrast, if we focus exclusively on research from 2021 onward, it becomes evident that deep learning models, namely LSTM, DNN (Deep Neural Networks) (Canale et al., 2021), and BERT, have outperformed others in the majority of recent studies.

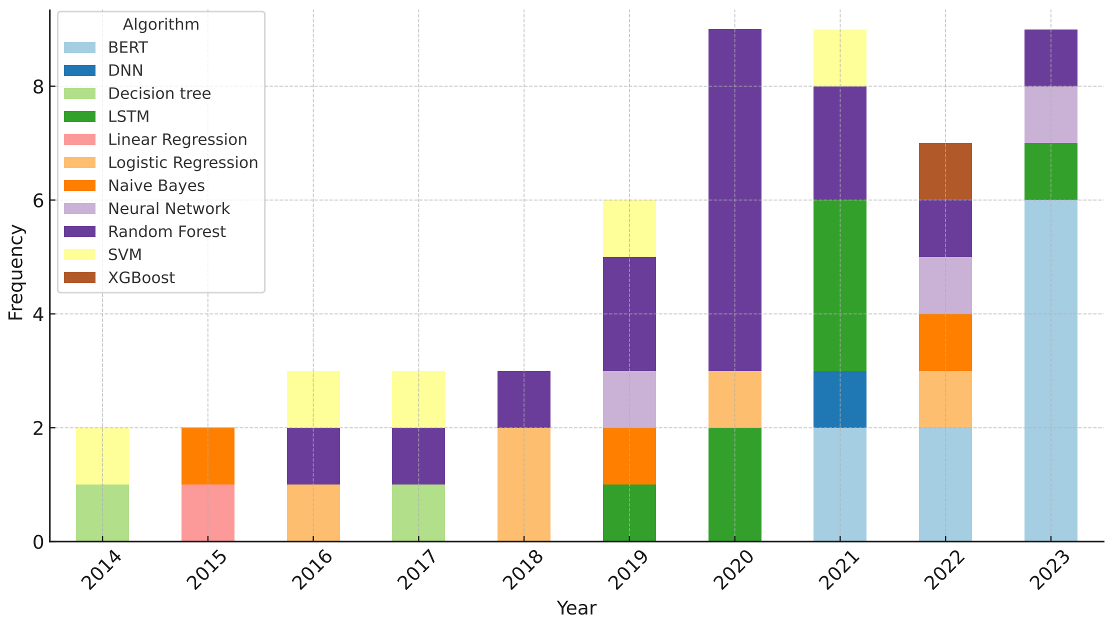


Figure 8. The highest-performance machine learning algorithms in the studies that used NLP for LA over the years.

In our final analysis, we aimed to thoroughly examine the outcomes from the chosen studies. For this purpose, we focused on adapting Cohen's kappa, largely used for its robustness in evaluating the quality of machine learning algorithms (A. Barbosa et al., 2021; Y. Hu et al., 2020; G. Barbosa et al., 2020; M. Ferreira et al., 2020; Ullmann, 2017). Additionally, we considered accuracy, as it emerged as the most frequently used metric across the selected papers ($N = 45$).

Table 4 presents the mean, median, minimum, and maximum values for kappa and accuracy in the studies that included them as part of the results divided by educational tasks and overall sum. In all cases, we can see that the mean and median values are very close. Additionally, no significant variation in mean and median was detected for the tasks presented (Assessment, Collaborative Learning, and Feedback) compared to the overall result. On the other hand, the tasks included in the "other" category (e.g., Reading Comprehension, Recommendation System, Student Modelling) reached lower values. It is important to highlight that other factors (e.g., dataset) could influence the outcomes, even for papers addressing similar tasks. Therefore, this result is not intended to serve as a definitive benchmark for the field but rather as an indication derived from this SLR.

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License (CC BY 4.0)

Table 4. Synthesis of selected studies regarding Cohen's kappa and accuracy metrics.

Educational Task	Measure	Number of Papers	Mean (SD)	Median	Min	Max
Assessment	kappa	12	0.57 (0.18)	0.56	0.29	0.95
	accuracy	15	0.82 (0.10)	0.83	0.58	0.98
Collaborative Learning	kappa	10	0.54 (0.20)	0.52	0.15	0.92
	accuracy	16	0.79 (0.11)	0.80	0.55	0.96
Feedback	kappa	05	0.60 (0.19)	0.59	0.28	0.95
	accuracy	09	0.81 (0.08)	0.81	0.67	0.97
Others	kappa	01	0.24 (0.00)	0.24	0.24	0.24
	accuracy	06	0.69 (0.13)	0.68	0.52	0.92
Overall	kappa	28	0.54 (0.19)	0.53	0.15	0.95
	accuracy	45	0.79 (0.10)	0.80	0.52	0.97

4. Discussion

The results of our literature review offer information in the three key dimensions: (i) comprehensive statistics that illustrate the evolution of the community, (ii) detailed information regarding the practical implementation of NLP algorithms in learning contexts, and (iii) a synthesis of technical aspects to indicate the primary standards and best practices for the integration of NLP within the LA domain.

Our results showed that the most active authors who worked on this topic are based at institutions in the US, the UK, Australia, and Canada, i.e., more than 59.3% of the papers reviewed were written by authors from these countries. This finding may explain why, in more than 80% of the studies we reviewed, the authors analyzed texts written in English. Further, the data reveals an upward trajectory in the volume of publications over time, culminating in a peak in 2023. This growth can also be associated with the overall expansion of the LA field.

Regarding the adoption of these methods in educational settings, our analysis highlights three critical aspects: the nature of tasks being addressed, the textual resources used for implementation, and the practical implications of these applications. First, the researchers appeared to focus on using text analytical approaches to understand and support collaborative learning, mainly in the applications related to online discussions. Studies identified in this group can be traced back to the original views on LA as follows (Siemens & Baker, 2012):

1. *Alignment between theory and experimentation/results:* The majority of the papers reviewed were grounded in theories, e.g., community of inquiry (Kovanović et al., 2016), knowledge building (M. A. D. Ferreira et al., 2022), and critical thinking (Wang et al., 2016);

2. *Concerns about the methodological process:* The papers we reviewed clearly describe the methodological procedures used. For instance, Farrow and colleagues (2019) elaborated on a series of steps performed to avoid data contamination in the analysis of online discussions.

3. *Focus on the learning process, not only on the outcome:* Although some papers were focused on predicting student performance, the papers related to online discussions tend to be focused more on understanding students' behaviour (Siemens & Baker, 2012).

4. *Application of methods for more visual and explainable insights:* The papers included in this review adopted a wide range of approaches, integrating NLP algorithms with several other analytical methods to enrich their findings. For instance, social network analysis (Lee & Tan, 2017; Boroujeni et al., 2017) was frequently used to explore relational patterns, while epistemic network analysis (M. Ferreira et al., 2020; M. A. D. Ferreira et al., 2022) provided a lens to examine the structure of knowledge and reasoning processes. Additionally, some studies employed a combination of these techniques (Swiecki & Shaffer, 2020), leveraging the strengths of both methods to generate deeper insights into the data, ultimately contributing to a more comprehensive understanding of the subject matter.

5. *Adoption of explainable white-box algorithms:* Many papers focused on using white-box machine learning algorithms to provide further information on the decision-making of the models.

A similar analysis can be applied to other significant tasks: assessment and feedback provision. In assessment, studies primarily concentrate on scoring essays (Allen et al., 2016; X. Hu, 2017; Ferreira Mello et al., 2022; Oliveira et al., 2023) and short-answer responses (Qiao & Hu, 2019; Erickson et al., 2020). On the other hand, the scope of feedback provision is much broader, encompassing a diverse array of contexts, including enhancing the effectiveness of teachers (or peers) in delivering feedback (Dood et al., 2022; Mello et al., 2022; Hunkins et al., 2022) and developing automated systems to provide students with pertinent and timely information (Knight et al., 2018; Del Gobbo et al., 2023). The community has attempted to incorporate the five items outlined in both scenarios.

However, the selected papers demonstrated the lack of practical applications of NLP. In other words, the community has been developing many algorithms and models that have not been evaluated in an actual educational application. Among all the selected papers, 89.10% evaluated the proposed approaches on previously collected data, and only 11 (out of 156) indicated some positive implications in learning outcomes or supported learning and/or teaching. One reason for this issue is the complexity involved with text analysis applications. Many studies focus on specific parts of a bigger problem (e.g., identification of rhetorical categories that could lead to an essay scoring system; Ferreira Mello et al., 2022). The studies have so far tried to solve fundamental problems of analysis of textual data, which is already a challenging task on its own, without considering aspects related to co-designing user interfaces that use insights of text analytics to support and enhance learning and teaching (Herodotou et al., 2020). As some areas of NLP in LA solidify (e.g., a large body of research on the classification of discussion messages according to the cognitive and social presence of the communities of inquiry framework), we can expect this transition to happen. However, we as a community need to recognize the need to develop a strong text analytic foundation and intentionally close the loop with insights obtained through NLP.

In our final contribution, we expect to establish a technical benchmark for future applications in NLP within LA. To achieve this goal, we have systematically encapsulated information about three fundamental aspects of an NLP study: the dataset used, the algorithms implemented, and the specific features analyzed. First, our review indicated that most of the selected papers performed experimentation with small to midsize datasets when compared to traditional NLP evaluation in general (Khurana et al., 2023; Min et al., 2023). In some cases, only 25 instances were adopted in the evaluation. Generally, the median sizes of datasets used in this context vary from 477 to 4,246 instances.

Our results also revealed that only five datasets were publicly available for other researchers (we investigated only the datasets described in the papers). In LA, this issue happens mainly due to ethical concerns. However, the artificial intelligence community has put much effort into delivering reproducibility in their studies. Furthermore, we previously noted a significant dependence on English-language datasets for evaluating the proposed models, with 115 out of the 156 selected papers exclusively using data in English for their experiments.

The RQ6 analysis indicates that decision tree algorithms, including Decision Trees, Random Forest, XGBoost, and AdaBoost, remain highly prevalent in the field (Iqbal et al., 2023; Oliveira et al., 2023; Ferreira Mello et al., 2022). This preference could be attributed to the transparency of their decision-making processes, which facilitates deeper qualitative insights. Conversely, there is a noticeable shift toward integrating state-of-the-art features and models in NLP applications in educational contexts. Since 2021, advanced deep learning architectures have been increasingly adopted, including word embeddings as features and LSTM (Ruan et al., 2021; Park et al., 2021; Guo et al., 2023) and BERT (Morris et al., 2023; Albano et al., 2023; Liu et al., 2023) as models. Recent research has begun exploring LLMs (Li et al., 2023; Qureshi, 2023; Rajala et al., 2023), such as ChatGPT, demonstrating promising results. However, the number of papers on this subject remains relatively modest, reflecting the novelty and emergent status of this technology.

Finally, RQ7 aimed to indicate the level of achievement of different NLP models for LA applications. The lowest outcomes retrieved in the studies analyzed were 0.15 in kappa and 0.52 in accuracy (A. Barbosa et al., 2021). However, the mean performance metrics observed in the field significantly exceed these thresholds, typically reaching 0.54 in kappa and 0.79 in accuracy. These figures are noteworthy: a kappa value of 0.54 denotes moderate agreement, indicating a reliable degree of consistency in the model's predictions compared to random chance (Cohen, 1960). Similarly, an accuracy of 0.79 is recognized as a strong indicator of good performance, reflecting a high rate of correct predictions by the model (Aggarwal & Zhai, 2012; Selva Birunda & Kanniga Devi, 2021).

In summary, our analysis indicates that the most commonly used aspects of examining NLP within the LA domain would involve (i) using a dataset comprising a minimum of 1,000 instances; (ii) combining the Random Forest algorithm with TF-IDF, Structural Features, LIWC, and Coh-Metrix as baseline models; and (iii) evaluating word embeddings through BERT and LSTM classifiers.

In conclusion, we highlight several critical considerations for the community involved in NLP for LA, identifying potential directions for future research. These recommendations are drawn from both the findings of this study and the broader body of literature on LA. As the field continues to evolve, it is essential to focus on these emerging directions to address current limitations, harness new opportunities, and foster innovation in applying NLP techniques to educational data analysis.

Replicability of the studies: One necessity to improve the community achievements is creating mechanisms to enhance the reproducibility of the studies (Kitto et al., 2023). It would include developing standard approaches and benchmark datasets, serving as valuable resources for research comparison and evaluation (Wilson et al., 2017). In this context, prioritizing research in privacy protection methods, such as de-identification and pseudonymization, is crucial (Farrow et al., 2023). These methods are instrumental in enabling the sharing of textual datasets across different research groups while safeguarding privacy (Tsai et al., 2020). In the cases where privacy concerns present an obstacle in data sharing, it is important to encourage researchers to make their programming codes available, thereby promoting transparency and progress in the field (Tsai et al., 2020).

Evaluation for different contexts: The datasets from the selected studies showed a distinct bias toward the English language, and many of them focused on a single domain (Ferreira-Mello et al., 2019). This observation highlights a pressing need to expand the scope of research to encompass additional educational fields and a variety of languages. Such diversification is essential for developing and evaluating classifiers that are representative and effective across different disciplines and linguistic contexts (Ahadi et al., 2022; Cavalcanti et al., 2023; A. Barbosa et al., 2021). Therefore, it is necessary to undertake further data collection efforts. These efforts should aim at gathering a more inclusive and diverse range of data, ensuring that future research is not only more comprehensive but also more universally applicable and relevant to a broader spectrum of educational and linguistic settings.

Use of LLMs: Although the SLR selected a few papers using LLMs, this continues to be an area to be explored in future studies (Yan et al., 2024). LLMs offer new possibilities to impact educational research and practice with tasks such as text analysis, language generation, and personalized learning. However, the community still does not have concrete evidence of the application of LLMs to address problems already established with other NLP algorithms. It is important to highlight that the ethical considerations and challenges related to the deployment of LLMs in educational settings, such as bias, privacy, and the impact on learning equity, should be rigorously examined.

Explainability: This literature review has emphasized the significance of employing white-box machine learning approaches; in many selected papers the authors used decision trees, regression, and naive Bayes algorithms. However, recent papers suggest an increasing integration of deep learning models (e.g., LSTM and BERT). Thus, it is important to adopt methods to further unpack the decisions made by these classifiers. One alternative is the use of explainable artificial intelligence (XAI) techniques for analyzing educational data. Initial studies indicate that combining deep learning with XAI enhances the outcome and brings clarity and understanding to the decision-making processes of artificial intelligence systems in educational contexts (Kumar & Boulanger, 2020; Oliveira et al., 2023; Cavalcanti et al., 2023).

Practical use: This SLR observed that most chosen publications primarily focus on proof-of-concept studies, analyzing past data. While these automated methods demonstrate effectiveness in controlled environments, their practical application could fall short. Future research should incorporate the proposed classifiers into real-world environments to bridge this gap. This integration will allow for a comprehensive evaluation of the system's intrinsic validity and its practical utility for students and instructors (Tsai & Gasevic, 2017; Márquez et al., 2024).

Measurement of student learning gains: Another relevant future direction in this field involves measuring student learning gains more effectively. The LA community has been discussing closing the loop and measuring learning gains in the last few years (Ferguson et al., 2023). Future studies should concentrate on developing and implementing analytical tools and methods that track and enhance student learning outcomes. By doing so, the field can evolve from experimental studies with data to actively contributing to improving educational practices and student performance.

5. Limitations

The primary limitation of this study arises from our search methodology, which specifically targeted papers using "learning analytics" as a keyword. This approach may have excluded relevant studies that do not explicitly mention this keyword, though it was chosen to focus our initial search phase. Second, due to space limitations, we could not explicitly include the references and the details of each paper selected in a spreadsheet. We published a website with all the information extracted from the papers[3]. Third, we acknowledge that many recent studies use LLMs as NLP models for various applications. However, this method was not extensively covered in our work because only a few of the selected articles (156) used LLMs. Nevertheless, we have discussed potential future directions for the field. Finally, it is important to mention that certain papers presented limited details regarding their methods and techniques, resulting in several studies being coded with labels such as "no details" and "no evidence" in various categories of our review. Despite this, we retained these papers in our analysis since they offered valuable insights for at least one RQ. However, it is crucial to note that we conducted a rigorous evaluation to ensure the overall quality and relevance of the papers included in our study.

[3]Available at https://shorturl.at/bMdv1

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

This study was partially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (310888/2021-2).

References

Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. In C. C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 163–222). Springer. https://doi.org/10.1007/978-1-4614-3223-4_6

Aguerrebere, C., Cobo, C., Gomez, M., & Mateu, M. (2017). Strategies for data and learning analytics informed national education policies: The case of Uruguay. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 449–453). ACM. https://doi.org//10.1145/3027385.3027444

Ahadi, A., Singh, A., Bower, M., & Garrett, M. (2022). Text mining in education—A bibliometrics-based systematic review. *Education Sciences*, *12*(3), 210. https://doi.org/10.3390/educsci12030210

Albano, V., Firmani, D., Laura, L., Mathew, J. G., Paoletti, A. L., & Torrente, I. (2023). NLP-based management of large multiple-choice test item repositories. *Journal of Learning Analytics*, *10*(3), 28–44. https://doi.org/10.18608/jla.2023.7897

Albano, V., Firmani, D., Laura, L., Paoletti, A. L., & Torrente, I. (2022). Managing large multiple-choice test items repositories. In E. Banissi, A. Ursyn, M. W. M. Bannatyne, J. M. Pires, N. Datia, K. Nazemi, B. Kovalerchuk, R. Andonie, M. Nakayama, F. Sciarrone, W. Huang, Q. V. Nguyen, M. S. Mabakane, A. Rusu, M. Temperini, U. Cvek, M. Trutschl, H. Mueller, H. Siirtola, . . . V. Geroimenko (Eds.), *Proceedings of the 26th International Conference on Information Visualisation* (IV 2022), 19–22 July 2022, Vienna, Austria (pp. 275–279). IEEE. https://doi.org/10.1109/IV56949.2022.00054

Allen, L. K., Mills, C., Jacovina, M. E., Crossley, S., D'Mello, S., & McNamara, D. S. (2016). Investigating boredom and engagement during writing using multiple sources of information: The essay, the writer, and keystrokes. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (LAK 2016), 25–29 April 2016, Edinburgh, UK (pp. 114–123). ACM. https://doi.org/10.1145/2883851.2883939

Allen, L. K., Mills, C., Perret, C., & McNamara, D. S. (2019). Are you talking to me? Multi-dimensional language analysis of explanations during reading. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 116–120). ACM. https://doi.org/10.1145/3303772.3303835

Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Are you reading my mind? Modeling students' reading comprehension skills with natural language processing techniques. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (LAK 2015), 16–20 March 2015, Poughkeepsie, New York, USA (pp. 246–254). ACM. https://doi.org/10.1145/2723576.2723617

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, *50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069

Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, *46*(4), 557–590. https://doi.org/10.1007/s11747-018-0581-x

Barbosa, A., Ferreira, M., Mello, R. F., Lins, R. D., & Gasevic, D. (2021). The impact of automatic text translation on classification of online discussions for social and cognitive presences. In *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK 2021), 12–16 April 2021, Irvine, California, USA (pp. 77–87). ACM. https://doi.org/10.1145/3448139.3448147

Barbosa, G., Camelo, R., Cavalcanti, A. P., Miranda, P., Mello, R. F., Kovanović, V., & Gašević, D. (2020). Towards automatic cross-language classification of cognitive presence in online discussions. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 605–614). ACM. https://doi.org/10.1145/3375462.3375496

Benedetto, L., Cappelli, A., Turrin, R., & Cremonesi, P. (2020). R2DE: A NLP approach to estimating IRT parameters of newly generated questions. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 412–421). ACM. https://doi.org/10.1145/3375462.3375517

Boroujeni, M. S., Hecking, T., Hoppe, H. U., & Dillenbourg, P. (2017). Dynamics of MOOC discussion forums. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 128–137). ACM. https://doi.org/10.1145/3027385.3027391

Buckingham Shum, S., Sándor, Á., Goldsmith, R., Wang, X., Bass, R., & McWilliams, M. (2016). Reflecting on reflective writing analytics: Assessment challenges and iterative evaluation of a prototype tool. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (LAK 2016), 25–29 April 2016, Edinburgh, UK (pp. 213–222). ACM. https://doi.org/10.1145/2883851.2883955

Canale, L., Farinetti, L., & Cagliero, L. (2021). From teaching books to educational videos and vice versa: A cross-media content retrieval experience. In W. K. Chan, B. Claycomb, H. Takakura, J.-J. Yang, Y. Teranishi, D. Towey, S. Segura, H. Shahriar, S. Reisman, & S. I. Ahamed (Eds.), *Proceedings of the IEEE 45th Annual Computers, Software, and Applications Conference* (COMPSAC 2021), 12–16 July 2021, online (pp. 115–120). IEEE. https://doi.org/10.1109/COMPSAC51774.2021.00027

Carnell, S., Lok, B., James, M. T., & Su, J. K. (2019). Predicting student success in communication skills learning scenarios with virtual humans. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 436–440). ACM. https://doi.org/10.1145/3303772.3303828

Castañeda, L., & Selwyn, N. (2018). More than tools? Making sense of the ongoing digitizations of higher education. *International Journal of Educational Technology in Higher Education*, *15*(1), 1–10. https://doi.org/10.1186/s41239-018-0109-y

Cavalcanti, A. P., Diego, A., Mello, R. F., Mangaroska, K., Nascimento, A., Freitas, F., & Gašević, D. (2020). How good is my feedback? A content analysis of written feedback. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 428–437). ACM. https://doi.org/10.1145/3375462.3375477

Cavalcanti, A. P., Mello, R. F., Gašević, D., & Freitas, F. (2023). Towards explainable prediction feedback messages using BERT. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-023-00375-w

Chang, X., Wang, B., & Hui, B. (2022). Towards an automatic approach for assessing program competencies. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 119–129). ACM. https://doi.org/10.1145/3506860.3506875

Chen, B., & Resendes, M. (2014). Uncovering what matters: Analyzing transitional relations among contribution types in knowledge-building discourse. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (LAK 2014), 24–28 March 2014, Indianapolis, Indiana, USA (pp. 226–230). ACM. https://doi.org/10.1145/2567574.2567606

Chen, G., Rolim, V., Mello, R. F., & Gašević, D. (2020). Let's shine together! A comparative study between learning analytics and educational data mining. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 544–553). ACM. https://doi.org/10.1145/3375462.3375500

Chowdhary, K. R. (2020). Natural language processing. In *Fundamentals of artificial intelligence* (pp. 603–649). Springer. https://doi.org/10.1007/978-81-322-3972-7_19

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cross, S., Waters, Z., Kitto, K., & Zuccon, G. (2017). Classifying help seeking behaviour in online communities. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 419–423). ACM. https://doi.org/10.1145/3027385.3027442

Crossley, S., Liu, R., & McNamara, D. (2017). Predicting math performance using natural language processing tools. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 339–347). ACM. https://doi.org/10.1145/3027385.3027399

Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (LAK 2016), 25–29 April 2016, Edinburgh, UK (pp. 6–14). ACM. https://doi.org/10.1145/2883851.2883931

Dascalu, M., Trausan-Matu, S., Dessus, P., & McNamara., D. S. (2015). Discourse cohesion: A signature of collaboration. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (LAK 2015), 16–20 March 2015, Poughkeepsie, New York, USA (pp. 350–354). ACM. https://doi.org/10.1145/2723576.2723578

Del Gobbo, E., Guarino, A., Cafarelli, B., & Grilli, L. (2023). GradeAid: A framework for automatic short answers grading in educational contexts—design, implementation and evaluation. *Knowledge and Information Systems*, *65*(10), 4295–4334. https://doi.org/10.1007/s10115-023-01892-9

Donnelly, P. J., Blanchard, N., Olney, A. M., Kelly, S., Nystrand, M., & D'Mello, S. K. (2017). Words matter: Automatic detection of teacher questions in live classroom discourse using linguistics, acoustics, and context. In *Proceedings* *of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 218–227). ACM. https://doi.org/10.1145/3027385.3027417

Dood, A., Winograd, B., Finkenstaedt-Quinn, S., Gere, A., & Shultz, G. (2022). PeerBERT: Automated characterization of peer review comments across courses. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 492–499). ACM. https://doi.org/10.1145/3506860.3506892

Erickson, J. A., Botelho, A. F., McAteer, S., Varatharaj, A., & Heffernan, N. T. (2020). The automated grading of student open responses in mathematics. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 615–624). ACM. https://doi.org/10.1145/3375462.3375523

Farrow, E., Moore, J., & Gašević, D. (2019). Analysing discussion forum data: A replication study avoiding data contamination. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 170–179). ACM. https://doi.org/10.1145/3303772.3303779

Farrow, E., Moore, J. D., & Gasevic, D. (2023). Names, nicknames, and spelling errors: Protecting participant identity in learning analytics of online discussions. In *Proceedings of the 13th International Conference on Learning Analytics and Knowledge* (LAK 2023), 13–17 March 2023, Arlington, Texas, USA (pp. 145–155). ACM. https://doi.org/10.1145/3576050.3576070

Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 56–65). ACM. https://doi.org/10.1145/3027385.3027396

Ferguson, R., Khosravi, H., Kovanović, V., Viberg, O., Aggarwal, A., Brinkhuis, M., Buckingham Shum, S., Chen, L. K., Drachsler, H., Guerrero, V. A., Hanses, M., Hayward, C., Hicks, B., Jivet, I., Kitto, K., Kizilcec, R., Lodge, J. M., Manly, C. A., Matz, R. L., . . . Yan, V. X. (2023). Aligning the goals of learning analytics with its research scholarship: An open peer commentary approach. *Journal of Learning Analytics*, *10*(2), 14–50. https://doi.org/10.18608/jla.2023.8197

Ferguson, R., & Shum, S. B. (2012). Social learning analytics: Five approaches. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK 2012), 29 April–2 May 2012, Vancouver, British Columbia, Canada (pp. 23–33). ACM. https://doi.org/10.1145/2330601.2330616

Ferreira, M., Rolim, V., Mello, R. F., Lins, R. D., Chen, G., & Gašević, D. (2020). Towards automatic content analysis of social presence in transcripts of online discussions. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 141–150). ACM. https://doi.org/10.1145/3375462.3375495

Ferreira, M. A. D., Mello, R. F., Kovanovic, V., Nascimento, A., Lins, R., & Gasevic, D. (2022). NASC: Network analytics to uncover socio-cognitive discourse of student roles. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 415–425). ACM. https://doi.org/10.1145/3506860.3506978

Ferreira Mello, R., Fiorentino, G., Oliveira, H., Miranda, P., Rakovic, M., & Gasevic, D. (2022). Towards automated content analysis of rhetorical structure of written essays using sequential content-independent features in Portuguese. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 404–414). ACM. https://doi.org/10.1145/3506860.3506977

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(6), e1332. https://doi.org/10.1002/widm.1332

Fiallos, A., & Ochoa, X. (2019). Semi-automatic generation of intelligent curricula to facilitate learning analytics. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 46–50). ACM. https://doi.org/10.1145/3303772.3303834

Fougt, S. S., Siebert-Evenstone, A., Eagan, B., Tabatabai, S., & Misfeldt, M. (2018). Epistemic network analysis of students' longer written assignments as formative/summative evaluation. In *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (LAK 2018), 7–9 March 2018, Sydney, New South Wales, Australia (pp. 126–130). ACM. https://doi.org/10.1145/3170358.3170414

Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., & Knight, S. (2017). Reflective writing analytics for actionable feedback. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 153–162). ACM. https://doi.org/10.1145/3027385.3027436

Guo, L., Du, J., & Zheng, Q. (2023). Understanding the evolution of cognitive engagement with interaction levels in online learning environments: Insights from learning analytics and epistemic network analysis. *Journal of Computer Assisted Learning*, *39*(3), 984–1001. https://doi.org/10.1111/jcal.12781

Herodotou, C., Rienties, B., Hlosta, M., Boroowa, A., Mangafa, C., & Zdrahal, Z. (2020). The scalable implementation of predictive learning analytics at a distance learning university: Insights from a longitudinal case study. *The Internet and Higher Education*, *45*, 100725. https://doi.org/10.1016/j.iheduc.2020.100725

Hsiao, I.-H., & Awasthi, P. (2015). Topic facet modeling: Semantic visual analytics for online discussion forums. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (LAK 2015), 16–20 March 2015, Poughkeepsie, New York, USA (pp. 231–235). ACM. https://doi.org/10.1145/2723576.2723613

Hu, X. (2017). Automated recognition of thinking orders in secondary school student writings. *Learning: Research and Practice*, *3*(1), 30–41. https://doi.org/10.1080/23735082.2017.1284253

Hu, Y., Donald, C., Giacaman, N., & Zhu, Z. (2020). Towards automated analysis of cognitive presence in MOOC discussions: A manual classification study. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 135–140). ACM. https://doi.org/10.1145/3375462.3375473

Hunkins, N., Kelly, S., & D'Mello, S. (2022). "Beautiful work, you're rock stars!": Teacher analytics to uncover discourse that supports or undermines student motivation, identity, and belonging in classrooms. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 230–238). ACM. https://doi.org/10.1145/3506860.3506896

Iqbal, S., Rakovic, M., Chen, G., Li, T., Ferreira Mello, R., Fan, Y., Fiorentino, G., Radi Aljohani, N., & Gasevic, D. (2023). Towards automated analysis of rhetorical categories in students essay writings using Bloom's taxonomy. In *Proceedings of the 13th International Conference on Learning Analytics and Knowledge* (LAK 2023), 13–17 March 2023, Arlington, Texas, USA (pp. 418–429). ACM. https://doi.org/10.1145/3576050.3576112

Iqbal, S., Swiecki, Z., Joksimovic, S., Ferreira Mello, R., Aljohani, N., Ul Hassan, S., & Gasevic, D. (2022). Uncovering associations between cognitive presence and speech acts: A network-based approach. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 315–325). ACM. https://doi.org/10.1145/3506860.3506908

Jayakodi, K., Bandara, M., Perera, I., & Meedeniya, D. (2016). WordNet and cosine similarity based classifier of exam questions using Bloom's taxonomy. *International Journal of Emerging Technologies in Learning (Online)*, *11*(4), 142. https://doi.org/10.3991/ijet.v11i04.5654

Jung, Y., & Friend Wise, A. (2020). How and how well do students reflect? Multi-dimensional automated reflection assessment in health professions education. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 595–604). ACM. https://doi.org/10.1145/3375462.3375528

Khalil, M., Prinsloo, P., & Slade, S. (2021). Realising the potential of learning analytics: Reflections from a pandemic. In J. Liebowitz (Ed.), *Online learning analytics* (pp. 79–94). Auerbach Publications. https://doi.org/10.1201/9781003194620-5

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia tools and applications*, *82*(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Kitchenham, B. A., Budgen, D., & Brereton, P. (2015). *Evidence-based software engineering and systematic reviews*. Chapman & Hall/CRC.

Kitto, K., Manly, C. A., Ferguson, R., & Poquet, O. (2023). Towards more replicable content analysis for learning analytics. In *Proceedings of the 13th International Conference on Learning Analytics and Knowledge* (LAK 2023), 13–17 March 2023, Arlington, Texas, USA (pp. 303–314). ACM. https://doi.org/10.1145/3576050.3576096

Knight, S., Buckingham Shum, S., Ryan, P., Sándor, Á., & Wang, X. (2018). Designing academic writing analytics for civil law student self-assessment. *International Journal of Artificial Intelligence in Education*, *28*, 1–28. https://doi.org/10.1007/s40593-016-0121-0

Knight, S., & Littleton, K. (2015). Developing a multiple-document-processing performance assessment for epistemic literacy. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (LAK 2015), 16–20 March 2015, Poughkeepsie, New York, USA (pp. 241–245). ACM. https://doi.org/10.1145/2723576.2723577

Kong, B., Hemberg, E., Bell, A., & O'Reilly, U.-M. (2023). Investigating student's problem-solving approaches in MOOCs using natural language processing. In *Proceedings of the 13th International Conference on Learning Analytics and Knowledge* (LAK 2023), 13–17 March 2023, Arlington, Texas, USA (pp. 262–272). ACM. https://doi.org/10.1145/3576050.3576091

Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., & Dawson, S. (2018). Understand students' self-reflections through learning analytics. In *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (LAK 2018), 7–9 March 2018, Sydney, New South Wales, Australia (pp. 389–398). ACM. https://doi.org/10.1145/3170358.3170374

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (LAK 2016), 25–29 April 2016, Edinburgh, UK (pp. 15–24). ACM. https://doi.org/10.1145/2883851.2883950

Kumar, V., & Boulanger, D. (2020). Explainable automated essay scoring: Deep learning really has pedagogical value. *Frontiers in Education*, *5*, 572367. https://doi.org/10.3389/feduc.2020.572367

Lárusson, J. A., & White, B. (2012). Monitoring student progress through their written "point of originality." In *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK 2012), 29 April–2 May 2012, Vancouver, British Columbia, Canada (pp. 212–221). ACM. https://doi.org/10.1145/2330601.2330653

Lee, A. V. Y., & Tan, S. C. (2017). Temporal analytics with discourse analysis: Tracing ideas and impact on communal discourse. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 120–127). ACM. https://doi.org/10.1145/3027385.3027386

Leeman-Munk, S. P., Wiebe, E. N., & Lester, J. C. (2014). Assessing elementary students' science competency with text analytics. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (LAK 2014), 24–28 March 2014, Indianapolis, Indiana, USA (pp. 143–147). ACM. https://doi.org/10.1145/2567574.2567620

Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., Lyons, K., Gašević, D., & Chen, G. (2023). Can large language models write reflectively. *Computers and Education: Artificial Intelligence*, *4*, 100140. https://doi.org/10.1016/j.caeai.2023.100140

Liu, Z., Kong, X., Chen, H., Liu, S., & Yang, Z. (2023). MOOC-BERT: Automatically identifying learner cognitive presence from MOOC discussion data. *IEEE Transactions on Learning Technologies*, *16*(4), 528–542. https://doi.org/10.1109/TLT.2023.3240715

Márquez, L., Henríquez, V., Chevreux, H., Scheihing, E., & Guerra, J. (2024). Adoption of learning analytics in higher education institutions: A systematic literature review. *British Journal of Educational Technology*, *55*(2), 439–459. https://doi.org/10.1111/bjet.13385

McAuley, J., O'Connor, A., & Lewis, D. (2012). Exploring reflection in online communities. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK 2012), 29 April–2 May 2012, Vancouver, British Columbia, Canada (pp. 102–110). ACM. https://doi.org/10.1145/2330601.2330630

McNamara, D. S., Allen, L. K., Crossley, S. A., Dascalu, M., & Perret, C. A. (2017). Natural language processing and learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (Vol. 93). Society for Learning Analytics Research. https://doi.org/10.18608/hla17

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Mello, R. F., Neto, R., Fiorentino, G., Alves, G., Arêdes, V., Silva, J. V. G. F., Falcão, T. P., & Gašević, D. (2022). Enhancing instructors' capability to assess open-response using natural language processing and learning analytics. In I. Hilliger, P. J. Muñoz-Merino, T. D. Laet, A. Ortega-Arranz, & T. Farrell (Eds.), *Educating for a new future: Making sense of technology-enhanced learning adoption (EC-TEL 2022)* (pp. 102–115). Springer. https://doi.org/10.1007/978-3-031-16290-9_8

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heintz, I., & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, *56*(2), 1–40. https://doi.org/10.1145/360594

Molenaar, I., & Chiu, M. M. (2015). Effects of sequences of socially regulated learning on group performance. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (LAK 2015), 16–20 March 2015, Poughkeepsie, New York, USA (pp. 236–240). ACM. https://doi.org/10.1145/2723576.2723586

Morris, W., Crossley, S., Holmes, L., & Trumbore, A. (2023). Using transformer language models to validate peer-assigned essay scores in massive open online courses (MOOCs). In *Proceedings of the 13th International Conference on Learning Analytics and Knowledge* (LAK 2023), 13–17 March 2023, Arlington, Texas, USA (pp. 315–323). ACM. https://doi.org/10.1145/3576050.3576098

Nicoll, S., Douglas, K., & Brinton, C. (2022). Giving feedback on feedback: An assessment of grader feedback construction on student performance. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 239–249). ACM. https://doi.org/10.1145/3506860.3506897

Niemann, K., Schmitz, H.-C., Kirschenmann, U., Wolpers, M., Schmidt, A., & Krones, T. (2012). Clustering by usage: Higher order co-occurrences of learning objects. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK 2012), 29 April–2 May 2012, Vancouver, British Columbia, Canada (pp. 238–247). ACM. https://doi.org/10.1145/2330601.2330659

Oliveira, H., Ferreira Mello, R., Barreiros Rosa, B. A., Rakovic, M., Miranda, P., Cordeiro, T., Isotani, S., Bittencourt, I., & Gasevic, D. (2023). Towards explainable prediction of essay cohesion in Portuguese and English. In *Proceedings of the 13th International Conference on Learning Analytics and Knowledge* (LAK 2023), 13–17 March 2023, Arlington, Texas, USA (pp. 509–519). ACM. https://doi.org/10.1145/3576050.3576152

Öncel, P., Flynn, L. E., Sonia, A. N., Barker, K. E., Lindsay, G. C., McClure, C. M., McNamara, D. S., & Allen, L. K. (2021). Automatic student writing evaluation: Investigating the impact of individual differences on source-based writing. In *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK 2021), 12–16 April 2021, Irvine, California, USA (pp. 620–625). ACM. https://doi.org/10.1145/3448139.3448207

Paredes, W. C., & Chung, K. S. K. (2012). Modelling learning & performance: A social networks perspective. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK 2012), 29 April–2 May 2012, Vancouver, British Columbia, Canada (pp. 34–42). ACM. https://doi.org/10.1145/2330601.2330617

Park, K., Sohn, H., Mott, B., Min, W., Saleh, A., Glazewski, K., Hmelo-Silver, C., & Lester, J. (2021). Detecting disruptive talk in student chat-based discussion within collaborative game-based learning environments. In *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK 2021), 12–16 April 2021, Irvine, California, USA (pp. 405–415). ACM. https://doi.org/10.1145/3448139.3448178

Pugh, S. L., Rao, A. R., Stewart, A. E., & D'Mello, S. K. (2022). Do speech-based collaboration analytics generalize across task contexts? In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 208–218). ACM. https://doi.org/10.1145/3506860.3506894

Qiao, C., & Hu, X. (2019). Measuring knowledge gaps in student responses by mining networked representations of texts. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 275–279). ACM. https://doi.org/10.1145/3303772.3303822

Qureshi, B. (2023). ChatGPT in computer science curriculum assessment: An analysis of its successes and shortcomings. In *Proceedings of the Ninth International Conference on e-Society, e-Learning and e-Technologies* (ICSLT 2023), 9–11 June 2023, Portsmouth, UK (pp. 7–13). ACM. https://doi.org/10.1145/3613944.3613946

Rajala, J., Hukkanen, J., Hartikainen, M., & Niemelä, P. (2023). \"Call me Kiran\"—ChatGPT as a tutoring chatbot in a computer science course. In *Proceedings of the 26th International Academic Mindtrek Conference* (Mindtrek 2023), 3–6 October 2023, Tampere, Finland (pp. 83–94). ACM. https://doi.org/10.1145/3616961.3616974

Rakovic, M., Fan, Y., Van Der Graaf, J., Singh, S., Kilgour, J., Lim, L., Moore, J., Bannert, M., Molenaar, I., & Gasevic, D. (2022). Using learner trace data to understand metacognitive processes in writing from multiple sources. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 130–141). ACM. https://doi.org/10.1145/3506860.3506876

Raković, M., Iqbal, S., Li, T., Fan, Y., Singh, S., Surendrannair, S., Kilgour, J., van Der Graaf, J., Lim, L., Molenaar, I., Bannert, M., Moore, J., & Gašević, D. (2023). Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-text writing task. *Journal of Computer Assisted Learning*, *39*(3), 703–718. https://doi.org/10.1111/jcal.12769

Raković, M., Winne, P. H., Marzouk, Z., & Chang, D. (2021). Automatic identification of knowledge-transforming content in argument essays developed from multiple sources. *Journal of Computer Assisted Learning*, *37*(4), 903–924. https://doi.org/10.1111/jcal.12531

Robinson, C., Yeomans, M., Reich, J., Hulleman, C., & Gehlbach, H. (2016). Forecasting student achievement in MOOCs with natural language processing. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (LAK 2016), 25–29 April 2016, Edinburgh, UK (pp. 383–387). ACM. https://doi.org/10.1145/2883851.2883932

Ruan, S., Wei, W., & Landay, J. (2021). Variational deep knowledge tracing for language learning. In *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK 2021), 12–16 April 2021, Irvine, California, USA (pp. 323–332). ACM. https://doi.org/10.1145/3448139.3448170

Rüdian, S., Dittmeyer, M., & Pinkwart, N. (2022). Challenges of using auto-correction tools for language learning. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK 2022), 21–25 March 2022, online (pp. 426–431). ACM. https://doi.org/10.1145/3506860.3506867

Schlippe, T., & Sawatzki, J. (2022). AI-based multilingual interactive exam preparation. In D. Guralnick, M. Auer, & A. Poce (Eds.), *Innovations in learning and technology for the workplace and higher education: Proceedings of 'The Learning Ideas Conference 2021,' Lecture notes in networks and systems* (pp. 396–408, Vol. 349). Springer. https://doi.org/10.1007/978-3-030-90677-1_38

Sekiya, T., Matsuda, Y., & Yamaguchi, K. (2015). Curriculum analysis of CS departments based on CS2013 by simplified, supervised LDA. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (LAK

2015), 16–20 March 2015, Poughkeepsie, New York, USA (pp. 330–339). ACM. https://doi.org/10.1145/2723576.2723594

Selva Birunda, S., & Kanniga Devi, R. (2021). A review on word embedding techniques for text classification. In J. Raj, A. Iliyasu, R. Bestak, & Z. Baig (Eds.), *Innovative data communication technologies and application: Proceedings of ICIDCA 2020, Lecture notes on data engineering and communications technologies* (pp. 267–281, Vol. 59). Springer. https://doi.org/10.1007/978-981-15-9651-3_23

Serrat, O. (2017). Social network analysis. In *Knowledge solutions* (pp. 39–43). Springer. https://doi.org/10.1007/978-981-10-0983-9_9

Settles, B., & Meeder, B. (2016). A trainable spaced repetition model for language learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Vol. 1: Long Papers), 7–12 August 2016, Berlin, Germany (pp. 1848–1858). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-1174

Sherin, B. (2012). Using computational methods to discover student science conceptions in interview data. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK 2012), 29 April–2 May 2012, Vancouver, British Columbia, Canada (pp. 188–197). ACM. https://doi.org/10.1145/2330601.2330649

Shibani, A., Knight, S., & Buckingham Shum, S. (2019). Contextualizable learning analytics design: A generic model and writing analytics evaluations. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 210–219). ACM. https://doi.org/10.1145/3303772.3303785

Shibani, A., Koh, E., Lai, V., & Shim, K. J. (2016). Analysis of teamwork dialogue: A data mining approach. In J. Joshi, G. Karypis, L. Liu, X. Hu, R. Ak, Y. Xia, W. Xu, A.-H. Sato, S. Rachuri, L. Ungar, P. S. Yu, R. Govindaraju, & T. Suzumura (Eds.), *2016 IEEE International Conference on Big Data (Big Data)*, 5–8 December 2016, Washington, DC, USA (pp. 4032–4034). IEEE. https://doi.org/10.1109/BigData.2016.7841100

Shusterman, E., Kim, H. G., Facciotti, M., Igo, M., Sripathi, K., Karger, D., Segal, A., & Gal, K. (2021). Seeding course forums using the teacher-in-the-loop. In *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK 2021), 12–16 April 2021, Irvine, California, USA (pp. 22–31). ACM. https://doi.org/10.1145/3448139.3448142

Siemens, G., & Baker, R. S. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK 2012), 29 April–2 May 2012, Vancouver, British Columbia, Canada (pp. 252–254). ACM. https://doi.org/10.1145/2330601.2330661

Siemens, G., & Gasevic, D. (2012). Guest editorial—Learning and knowledge analytics. *Journal of Educational Technology & Society*, *15*(3), 1–2. https://drive.google.com/file/d/1SJQZSFOrix9_WZTvBtzvUL70bsLa_eqQ/view

Snow, E. L., Allen, L. K., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). You've got style: Detecting writing flexibility across time. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (LAK 2015), 16–20 March 2015, Poughkeepsie, New York, USA (pp. 194–202). ACM. https://doi.org/10.1145/2723576.2723592

Sousa, E., Alexandre, B., Ferreira Mello, R., Pontual Falcão, T., Vesin, B., & Gašević, D. (2021). Applications of learning analytics in high schools: A systematic literature review. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.737891

Southavilay, V., Yacef, K., Reimann, P., & Calvo, R. A. (2013). Analysis of collaborative writing processes using revision maps and probabilistic topic models. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK 2013), 8–13 April 2013, Leuven, Belgium (pp. 38–47). ACM. https://doi.org/10.1145/2460296.2460307

Stone, C., Quirk, A., Gardener, M., Hutt, S., Duckworth, A. L., & D'Mello, S. K. (2019). Language as thought: Using natural language processing to model noncognitive traits that predict college success. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 320–329). ACM. https://doi.org/10.1145/3303772.3303801

Swiecki, Z., & Shaffer, D. W. (2020). Isens: An integrated approach to combining epistemic and social network analyses. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 305–313). ACM. https://doi.org/10.1145/3375462.3375505

Tan, J. P.-L., Yang, S., Koh, E., & Jonathan, C. (2016, April). Fostering 21st century literacies through a collaborative critical reading and learning analytics environment: User-perceived benefits and problematics. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (LAK 2016), 25–29 April 2016, Edinburgh, UK (pp. 430–434). ACM. https://doi.org/10.1145/2883851.2883965

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Teplovs, C., Fujita, N., & Vatrapu, R. (2011). Generating predictive models of learner community dynamics. In *Proceedings of the First International Conference on Learning Analytics and Knowledge* (LAK 2011), 27 February–1 March 2011, Banff, Alberta, Canada (pp. 147–152). ACM. https://doi.org/10.1145/2090116.2090139

Thomas, D. A. (2014). Searching for significance in unstructured data: Text mining with Leximancer. *European Educational Research Journal*, *13*(2), 235–256. https://doi.org/10.2304/eerj.2014.13.2.235

Tsai, Y.-S., & Gasevic, D. (2017). Learning analytics in higher education—challenges and policies: A review of eight learning analytics policies. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 233–242). ACM. https://doi.org/10.1145/3027385.3027400

Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2020). The privacy paradox and its implications for learning analytics. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 230–239). ACM. https://doi.org/10.1145/3375462.3375536

Ullmann, T. D. (2017). Reflective writing analytics: Empirically determined keywords of written reflection. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, British Columbia, Canada (pp. 163–167). ACM. https://doi.org/10.1145/3027385.3027394

Vytasek, J. M., Patzak, A., & Winne, P. H. (2019). Topic development to support revision feedback. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, Arizona, USA (pp. 220–224). ACM. https://doi.org/10.1145/3303772.3303816

Wang, X., Wen, M., & Rosé, C. (2016). Towards triggering higher-order thinking behaviors in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (LAK 2016), 25–29 April 2016, Edinburgh, UK (pp. 398–407). ACM. https://doi.org/10.1145/2883851.2883964

Whitelock, D., Twiner, A., Richardson, J. T. E., Field, D., & Pulman, S. (2015). OpenEssayist: A supply and demand learning analytics tool for drafting academic essays. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (LAK 2015), 16–20 March 2015, Poughkeepsie, New York, USA (pp. 208–212). ACM. https://doi.org/10.1145/2723576.2723599

Wilson, A., Watson, C., Thompson, T. L., Drew, V., & Doyle, S. (2017). Learning analytics: Challenges and limitations. *Teaching in Higher Education*, *22*(8), 991–1007. https://doi.org/10.1080/13562517.2017.1332026

Yan, L., Martinez-Maldonado, R., & Gasevic, D. (2024). Generative artificial intelligence in learning analytics: Contextualising opportunities and challenges through the learning analytics cycle. In *Proceedings of the 14th International Conference on Learning Analytics and Knowledge* (LAK 2024), 18–22 March 2024, Kyoto, Japan (pp. 101–111). ACM. https://doi.org/10.1145/3636555.3636856

Zylich, B., & Lan, A. (2021). Linguistic skill modeling for second language acquisition. In *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK 2021), 12–16 April 2021, Irvine, California, USA (pp. 141–150). ACM. https://doi.org/10.1145/3448139.3448153