Journal of Computing in Higher Education
https://doi.org/10.1007/s12528-023-09370-5

Sentiment analysis for formative assessment in higher education: a systematic literature review

Carme Grimalt-Álvaro[1] · **Mireia Usart**[1]

Accepted: 23 March 2023

Abstract

Sentiment Analysis (SA), a technique based on applying artificial intelligence to analyze textual data in natural language, can help to characterize interactions between students and teachers and improve learning through timely, personalized feedback, but its use in education is still scarce. This systematic literature review explores how SA has been applied for learning assessment in online and hybrid learning contexts in higher education. Findings from this review show that there is a growing field of research on SA, although most of the papers are written from a technical perspective and published in journals related to digital technologies. Even though there are solutions involving different SA techniques that can help predicting learning performance, enhancing feedback and giving teachers visual tools, its educational applications and usability are still limited. The analysis evidence that the inclusion of variables that can affect participants' different sentiment expression, such as gender or cultural context, remains understudied and should need to be considered in future developments.

Keywords Artificial intelligence · Review of literature · Technology · Higher education · Gender

Introduction

In the last decade, there has been a significant raise in the amount of research articles, projects, and initiatives around the use of Big Data and artificial intelligence in education (Kastrati et al., 2021). Some of these studies have been developed in higher education, including online learning and hybrid or blended learning contexts. In these educational settings Virtual Learning Environments (VLE) such as Moodle are very common. Since the COVID19 pandemic there is

✉ Carme Grimalt-Álvaro
carme.grimalt@urv.cat

[1] Departament de Pedagogia, Facultat de Ciències de l'Educació i Psicologia, Universitat Rovira i Virgili, Tarragona, Spain

Published online: 18 April 2023

an extensive number of students' data stored and exchanged through these educational platforms.

With the aim of understanding and optimizing learning, different techniques for data analysis, measurement, collection, and reporting about learners' variables (e.g., time spent using the VLE, number of messages exchanged or type of interactions) are developed and grouped under the umbrella of Learning Analytics (LA) (Buckingham Shum & Ferguson, 2012). These techniques have been extensively studied in higher education in relation to performance and dropout predictions (Iglesias i Estradé, 2019), and assessment (Kastrati et al., 2021).

Sentiment Analysis (SA) is one of these LA techniques, defined by Mite-Baidal et al. (2018) as the contextual mining of unstructured text from documents, so that structured and insightful knowledge can be obtained and used for different goals. From the analysis of how students express themselves, especially in asynchronous environments, such as texts, forums, wikis or debates, teachers and students can have more information about students' satisfaction and reactions within a learning activity. Hence, teachers can appropriately and timely adapt their teaching, improving the quality of learning processes, while students can develop self-reflection. However, most of the SA techniques demand a high level of programming skills to be implemented (Kastrati et al., 2021). Additionally, they provide complex results in return, which hampers its use for almost all teachers and students. Therefore, it becomes necessary to explore which techniques have been developed, in which educational contexts, and also which of them integrates frontend modules or visual software to facilitate the interpretation of SA results among higher education teachers and students.

On the other hand, research on human communication has identified the existence of gendered patterns, which can be different for specific areas. For instance, Çoban et al. (2021) showed that male users on social media posted more positive messages as they were older, while the opposite was observed for females. Unfortunately, these patterns seem not having been integrated in the big data techniques. Thelwall (2018), argues that these techniques (and SA as well) can be gender biased. To promote a fair and inclusive education, it becomes necessary to explore how SA techniques can be more gender sensitive. It becomes necessary to study systematically previous contributions to have a better perspective on how SA can be applied inclusively in higher education.

This systematic literature review explores how SA has been implemented as a tool for assessment in online higher education research, to identify useful tools and techniques and if these have been applied within a gender perspective. To this end, this paper presents a brief literature review of the main conceptual points, the aims and research questions of the systematic literature review, and the methodology, based on the PRISMA statements (Moher et al., 2009). Results and discussion section shares the outcomes of the systematic literature review and raise different challenges identified from the analysis. Conclusions highlight the major contributions of our work, considering its limitations and future implications.

Literature review

SA for learning assessment in online higher education

Assessment of learning involves a process of verification, evaluation, and decision-making with the purpose of optimizing the teaching–learning process (Pinger et al., 2018). Formative and final assessment can increase motivation and involvement of students and provide opportunities for the correction of errors (Gikandi et al., 2011). Formative assessment represents a learning experience itself, developing students' autonomy, communication, and self-reflection and, in consequence, improving academic achievement (Martínez Cámara et al., 2016). An adequate formative assessment becomes essential for higher education online learning environments.

Different formats for providing formative assessment in online learning environments can be used, such a reflection papers/diaries, quizzes, wikis, discussion forums, blogs and e-portfolios (McLaughlin & Yan, 2017; Vonderwell & Boboc, 2013). Among these, discussion forums are the most common formative assessment tools (Xiong & Suen, 2018).

Traditionally, in online higher education, teachers carry out formative assessment from a one-to-one perspective. However, there is a consensus in the pedagogy literature that individuals learn from student–teacher interactions, but also from other social, student–student interactions (Borokhovski et al., 2012). From this perspective, different studies have investigated how online formative assessment can also integrate all these interactions and relate them to successful learning outcomes (Onan, 2021).

In the learning process, it is of particular interest how the affective domain may influence and be influenced by the interaction with peers. According to Buckingham Shum and Ferguson (2012), online learners may use the affective domain of learning to clarify their intentions, ground their learning, and engage in learning conversations. Kashy-Rosenbaum et al. (2018) measured significantly higher academic achievement within classrooms characterized by positive emotional environment, and significantly lower within classrooms characterized by negative emotional environment. In VLE, participants can share additional insights regarding course topics and provide their impressions and affective states through online forums and debates (Moreno-Marcos et al., 2019), easing the data storage of such information.

However, the enormous amount of information exchanged through VLE can exceed an affordable workload for teachers (McCarthy, 2017). This difficulty in digesting a big amount of textual data in a relatively short length of time may seriously jeopardize the understanding of social online learning and, in consequence, the quality of periodic formative assessment provided (McLaughlin & Yan, 2017).

For these reasons, in educational settings, different SA techniques have been developed as a proxy measure in real time students' sentiments (Yadegaridehkordi et al., 2019). SA techniques have been mostly used to improve the understanding of educational processes, study participants' satisfaction, and make performance and dropout predictions (Iglesias i Estradé, 2019). These techniques are designed and trained with a collection of textual data, and its application usually involves a high level of programming skills. As well, some researchers are starting to develop complementary front-end solutions to ease the technique management and interpreting results, such as edX-CAS or RAMS (Cobos et al., 2019; Elia et al., 2019), but there is not a global picture about what different solutions can be found in the literature and which of them are best suited for HE. Therefore, to leverage the benefits of formative assessment using SA, it becomes necessary to know what automated tools exist to help teachers to efficiently process and understand data from students' sentiments, so they can enhance the timely assessment of learning.

Gender Bias in SA, higher education and assessment

Unfortunately, SA techniques have paid little attention to potential differences in communication related of personal characteristics of participants, such as cultural differences, language barriers, age, and gender (Yadegaridehkordi et al., 2019). However, researchers have identified the existence of gendered patterns of communication between participants' uses of virtual environments. In VLE, females usually interact more than their male peers (Oreski & Kadoic, 2018; Van Horne et al., 2018), they usually participate to a lesser extent in the proposed activities with contributions that integrate fewer mistakes (Kickmeier-Rust et al., 2014). Moreover, Shapiro et al. (2017) observed that females expressed more negative views about their progress and self-perceived evaluation. These differences in communication patterns have not been considered in SA techniques development. This can jeopardize the effectiveness of SA as a generalized assessment tool, but also introduce bias and inequality in formative assessment.

Previous systematic literature analysis about SA

To establish the present context for this review, considering the prior research is mandatory. Relevant literature on the field of SA in education was examined, and three previous systematic literature reviews about SA were found: Kastrati et al. (2021), Mite-Baidal et al. (2018), and Zhou and Ye (2020). Mite-Baidal et al. (2018) studied the application of SA in e-learning in higher education. These authors focused their analysis on the resources and techniques used in relation to different SA analytical approaches, and the major benefits of SA. These benefits were "learning process improvement, performance improvement, reduction in course abandonment, teaching process improvement, and satisfaction with a course" (Mite-Baidal et al., 2018, p. 292), evidencing that the literature analysis had been carried out from the perspective of the improvement of the design of e-learning systems and teaching efficiency. The second review, carried out by Zhou and Ye (2020), followed a similar approach, but unlike the previous literature review, these researchers identified SA could also be used to help students perceive their emotions through the use of visual presentation and feedback.

However, apart from the improved ability to perceive and adjust one's emotions, the authors did not discuss how SA could be a tool to ameliorate students' formative assessment strategy through the improvement of self-reflective skills. Finally, Kastrati et al. (2021), carried out a more technical approach review, focused on the analysis of the datasets, solutions and the emotional expression and detection of the SA techniques used; they did not consider how SA techniques are used within the teaching—learning process whatsoever.

Previous systematic literature reviews have contributed to understand which SA techniques and its features have been used for educational purposes as well as, which applications for teaching and course design improvements can be identified in recent years. They evidence that the solutions are usually too technical, and that there is a lack of research effort on the use of general-purpose visualizing solutions (Kastrati et al., 2021). These reviews also identify a lack of research about how SA can be a tool for leveraging the quality of the formative assessment provided and how students' self-reflective skills and metacognitive abilities can be developed; as Treceñe (2019) highlights, and as well how SA can be use with a more inclusive approach. Therefore, this article focuses on the analysis of the use of SA as a tool to improve students' assessment and their learning processes from a gender-perspective approach, as a step in the development of fair SA techniques based on previous research evidence.

Aim and research questions

The body of literature is lacking a review that systematically classifies the research and results of the application of SA in higher education domain, related to students' feedback and assessment, considering possible gender differences in students' communication, and offering practical solutions for teachers. Hence, this literature review aims at exploring how Sentiment Analysis (SA) has been applied in higher education to identify best strategies in fostering gender-inclusive online learning environments. In particular, the following questions and sub-questions arise:

- (RQ1) Which are the different SA techniques used in formative assessment in higher education?
- (RQ2) Which are the different frontend software based on SA developed to provide teachers and students with timely and easy-to-understand SA analysis?
- (RQ3) Which are the general contributions of SA as a tool for students' assessment in higher education?
    - o (RQ3.1) From these general contributions, how has SA been used in higher education, from a gender perspective?

Methodology

This literature review was conducted in the following three review phases based on Kitchenham (2004): planning, conducting and reporting. Since this method requires an established search protocol and rigorous criteria for the screening and selection of the relevant publications, we utilized the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, as indicated in Liberati et al. (2009). These guidelines were used for the description of the eligibility criteria, information sources, data collection process, data items, and synthesis of results, as summarized in Fig. 1.

Study search strategy

The following search string was used to search in ERIC, Scopus, and Web of Science databases:

("distance education" OR "distance learning" OR "virtual learn*" OR "virtual classroom" OR "online learning" OR "Web-based learning" OR "electronic learning" OR blended OR hybrid OR "computer-assisted instruction") AND ("sentiment*" OR "sentiment analysis" OR "learning analytic*") AND ("assessment" OR "feedback*" OR "evaluation*" OR "formative assessment").

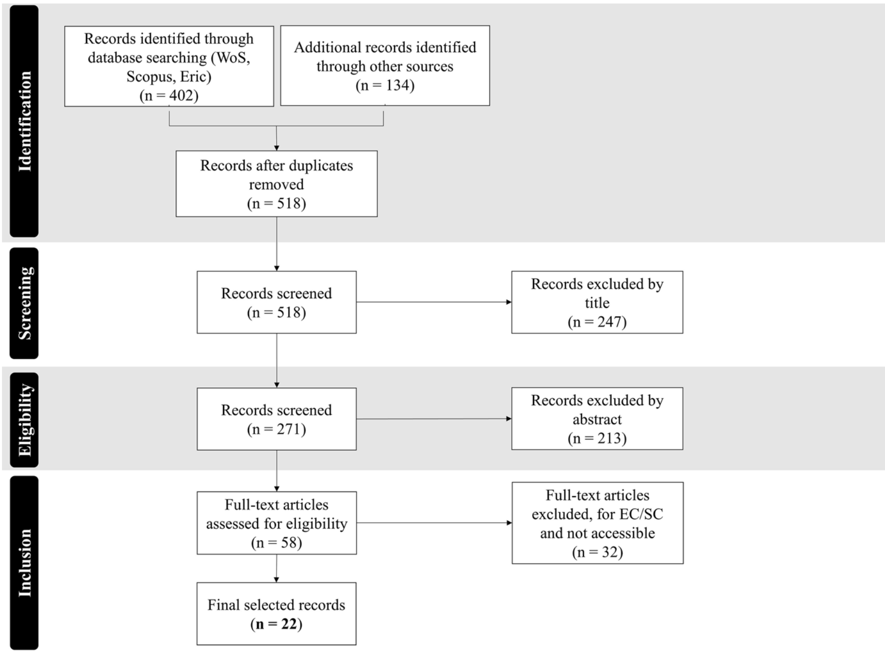

Fig. 1 Graphic summary of the literature review procedure, based on the PRISMA statement Liberati et al. (2009)

Studies considered for the identification phase could come from publicly available scholarly articles, book chapters, technical reports, dissertations, or presentations at scholarly meetings from Educational Sciences disciplines. We restricted the search to study title, abstract and keywords.

As results from the first strategy search were mostly from computer-science disciplines, manual searches were performed in 21 educational research journals, those were identified from the first strategy search, and also from the references in the three SA reviews mentioned in the introduction. Journals' websites were visited, and the same search string was used to search for articles related to higher education and teacher training as a complementary search.

Inclusion and exclusion criteria

Table 1 shows inclusion and exclusion criteria used in the screening, eligibility, and inclusion phases. All studies that met one or more exclusion criteria were excluded. Timeframe for publication's date was from January 2006 to June 2021. The 2006 year is considered to be the beginning of the "learning analytics age", according to Picciano (2012). The end date was set on June, as this study was conducted between May and August 2021.

Conducting the review

Before starting the review, the authors discussed and agreed on the research questions related to the information to be extracted from the publications. From this, the authors defined the search strategy and the inclusion and exclusion criteria.

Identification phase

In this phase, the documents from all the databases were retrieved using the search string. 402 publications were retrieved from the three databases and 134 publications were added from additional outstanding journals, identifying a total of 536 potential publications. We extracted information about the publication itself (i.e., title, authors and year of publication), learning outcomes, SA techniques and tools or software, and gender analysis, and incorporated it into a database for further analysis. This process identified and removed 14 duplicates and select 518 publications for the screening phase.

Screening phase

In the screening phase, documents were excluded by title applying the exclusion criteria (Table 1). We filtered 247 publications that were not meeting at least one exclusion criterion. A total of 271 publications were selected for the eligibility phase.

Table 1 Inclusion and exclusion criteria

Exclusion criteria	Selection criteria
EC1: The context of the publication was not pertinent (e.g., the study examined primary or secondary education or informal learning like social networks)	SC1: The publication reports pre-service or in-service teacher training or higher education learning contexts (universities)
EC2: The publication is not a primary study, it is a literature review, synthesis that compares and contrasts studies by other researchers	SC2: The publication reports a study in blended or online modalities (e.g., using a VLE like Moodle)
EC3: The publication is not in English	SC3: The publication included may also refer to the assessment of learners or providing feedback to learners with SA (not focusing on teachers or institution)
EC4: The publication is not in a correct format (e.g., the publication is a book review)	SC4: The publication provides concrete details on the SA technique (i.e., regarding the application of natural language processing (NLP), deep learning (DL), and machine learning (ML) as solutions for sentiment analysis in the education domain)
EC5: Full text not available	SC5: The publication may not present research results, but it describes a SA technique, or a frontend tool related to assessment of learning
EC6: The publication refers to LA but is not related to SA or to textual data analysis or assessment by teachers or students (e.g., Evaluates course policies or considers only students' logs)	

Eligibility Phase

Abstracts of the 271 publications were read to assess whether they satisfied the inclusion and exclusion criteria (Table 1). We excluded 213 publications, mostly because they did not focus on education (111), did not focus on higher education (70), or although the research context was educational, they focused only on technical aspects of SA (32). 58 publications were selected for the eligibility phase.

Inclusion phase

The 58 remaining publications were examined by reading them and applying the inclusion and exclusion criteria (Table 1). Finally, 31 publications were excluded. 4 of them were not accessible, and 27 publications did not meet the following inclusion criteria: SC1 (7), SC2 (6), SC 3 (12), and SC4 (2). We applied SC5 criteria to include references but still, we kept those ones not meeting it, to have a broader sample of final publications considered. 22 publications were finally selected. These publications were studied, based on the information in the database, by undertaking a qualitative synthesis to answer the research questions. Results are presented in the next section, following each of the three research questions.

Results and discussion

Dataset characteristics

A sample of 22 studies met all the IC for RQ 1,2, and 3. A first analysis of the search results evidence that although SA for evaluation in higher education is receiving increased attention by educational and technological research fields year by year (as

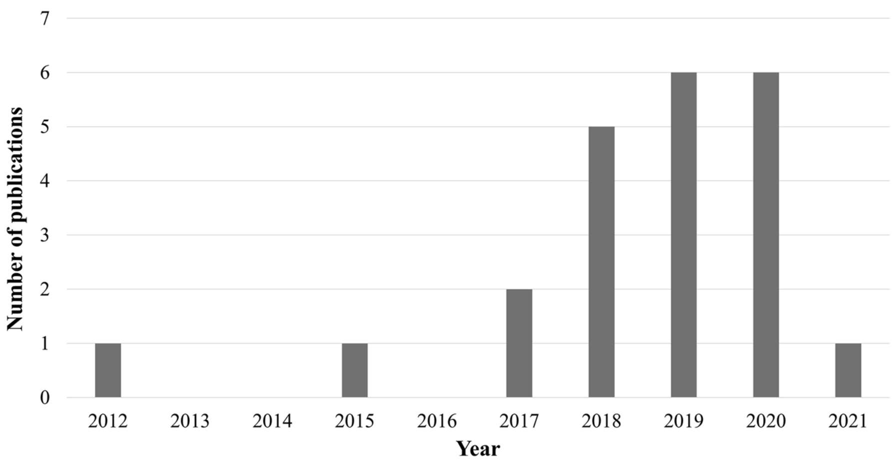

Fig. 2 Evolution of the number of references in the last 9 years. *Note*. Results of 2021 were included until June

can be seen in Fig. 2), it is still an emerging field. Hence, 11 of 22 studies were conference publications, which could be interpreted as a first step in the development before publishing in journals.

Regarding the origin of the publications, most of the references came from the American continent (eight references): USA (5), México (2) and Brazil (1). Six references were European: Greece (2), Spain (1), Portugal (1), Italy (1), and Finland (1). Five references were African: Nigeria (1), South Africa (1), Saudi Arabia (1), Morocco (1), and Egypt (1). Only two references were Asian: China (1) and Taiwan (1), and only one reference was from Australia.

Most of the studies focused on computer sciences (9 references; 41%), while other studies focused on science (3), engineering (1), and teacher training (2). Seven references did not specify the area of study, but mentioned undergraduate (2), postgraduate students (1) and learning communities (4). Moreover, in 11 studies (50%) authors used SA as an assessment tool in hybrid learning contexts and 11 (50%) in online learning. These results show that SA researchers usually came from STE(M) areas and from higher educational contexts where online learning plays an important role.

Regarding methodology, 14 studies were quantitative (from big-data analysis to longitudinal studies), while 4 were qualitative, and 4 used a mixed-methods approach. Because of this diversity in the methodological approach, samples sizes varied from 28 to 2600 direct participants, and when using statements or sentences as a sample, this ranged from 120 to 171,430 statements.

A first analysis evidences how research papers have evolved; while the first studies mainly focused on the methodological part (computer sciences perspective), up today studies consider educative application for higher education. The order of research questions tries to facilitate the presentation of major contributions and, for this reason, emergent themes that resulted from the analysis of the 22 publications considered, are described below.

RQ1. SA techniques used in formative assessment in higher education

Major contributions to answer RQ1 are structured around three topics; automatic data collection, information extraction and sentiment measurement, as follows:

Automatic data collection

There is an evolution towards the automatization of the previous stage of SA, referring to how data is collected and preprocessed before conducting SA. Major changes are identified specifically in the data collection gathered manually in the first studies, as in Abdulsalami et al. (2017) and Zhang et al. (2012). Different authors evidence the lack of a suitable integration between educational platforms and SA processing modules, becoming information retrieving a tedious task for teachers. To foster the adoption of SA in higher education, recent studies have worked on more complex systems integrating agents to collect posted students messages from diverse discussion forums or considered sources, as in Alencar and Netto (2020). Within this trend, future studies might develop methods to integrate students' productions on different supports beside text (Elia et al., 2019). The ethical implications that could emerge from a massive and systematic collection of students' productions need to be considered, especially regarding the transparency of institutions, the re-purposed data analysis, and the meaningful alternatives that participants have when not allowing their data to be shared (Andreotta et al., 2021).

Information extraction

Forum messages and emails become the main data sources on learning assessment in online Higher Education (Alencar & Netto, 2020; Camacho & Goel, 2018; Chaabi et al., 2019; Le et al., 2018; Osorio Angel et al., 2020). In blended scenarios, other data sources as video transcriptions and readings have been used (Cobos et al., 2019). Two major method approaches to SA are used to detect potential sentiments and opinions in texts: the lexicon-based approach, also called Natural Language Processing (NLP), and the Machine Learning (ML) approach. The most frequent approach is NLP (ten studies), consisting of the extraction of meaning from human language using previous libraries (lexicons) as a reference. NLP does not need to take context into account, and it uses an amount of open and free linguistic libraries available online (Zhang et al., 2012). It also enables a broader scope for making inferences, finding patterns in textual data, and inferring an emotion (Mostafa, 2020) with a high precision but a low recall. Hence, although the setting of the SA technique can be more accessible to develop, the lack of discipline-specific libraries is an important limitation, as some specialized terms can be wrongly associated with emotional states (Nunez, 2020). For this reason, some authors argue that using only a lexicon-based approach is not the best method of sentiment analysis, since important information may be missed (Alblawi & Alhamed, 2017).

The ML approach, used in five of the 22 references, is a subset of Artificial Intelligence (AI), in which the algorithm progressively "learns" to identify positive and negative sentiments, allowing a more accurate pattern prediction than NLP. However, ML needs supervised analysis and a higher amount of data to train the AI and make better predictions (Alencar & Netto, 2020), which is the main reason for its less frequent use. However, from our perspective, ML approaches might be more respectful of cultural diversity in Higher Education institutions, as not only would be more feasible to be implemented in institutions using a regional or minority language as vehicular, but also, respect the cultural differences in how students from same language speaking countries express themselves (e.g., Latin-American students vs. Spanish students). The remaining seven studies use both techniques (NLP and ML) to build a more solid process of SA, as both methods are non-exclusive and can be complementary used, as in Alblawi and Alhamed (2017).

Among these studies, an increasing trend of using third-parties SA software emerges. MeaningCloud (Bilro et al., 2022) or RStudio packages (Okoye et al., 2020) facilitate the utilization of SA in broader educational contexts and better explore its potential contributions in this field. However, we caution against the massive adoption of these solutions, as they might not be successfully enough in achieving current challenges of SA (e.g., use of specialized vocabulary, interpretation of negations and idioms, conducting SA at topic level within a text… in highly formal contexts), causing significant methodological flaws in research results.

Measuring sentiment

A final consideration is needed regarding the epistemological grounding of SA techniques. The most frequent approach in SA is to classify sentiments based on the measurement of the polarity or valence association (fourteen studies). Three studies measure sentiment polarity in a binary form (positive or negative), as in Abdulsalami et al. (2017) and Elia et al. (2019), eight studies classify sentiments at a three-level scale as in Camacho and Goel (2018) and Cobos et al. (2019), and three studies use polarity scales with over three levels, as in Alblawi and Alhamed (2017) and Gkontzis et al. (2020). There is an apparent consensus in the output in the SA analysis as a three-level polarity scale. However, although the use of positive, negative, or neutral scale has been reported to be the most frequent association within three-level polarity scales (Chiarello et al., 2020), a high dispersion in the measurements emerges. For example, by measuring extremely positive, positive, and non-positive sentiments (Zhang et al., 2012) or by using normalized compound between − 1 (extreme negative) and 1 (extreme positive) in Camacho and Goel (2018), Dehbozorgi et al. (2020), and Okoye et al. (2020).

Several studies relate sentiment polarity with students' motivation: learners with a positive attitude are more confident and motivated to learn (Mostafa, 2020; Weston et al., 2020 in Okoye et al., 2020). However, we argue for a more grounded interpretation of these SA results. The vast majority of publications perform SA at word or sentence level (i.e., attributing a sentiment score to each word or sentence and then computing the overall result), but motivation is a particular mental state, directed towards a particular topic, and usually related to behavioral responses. For this reason, we argue that measuring sentiment polarity provides information about the emotional climate of a group or general emotional state of a student (Usart et al., 2022), rather than providing information about a particular mental or emotional state. Hence, attributing polarity results to a mental state may carry the risk of falling into spurious relationships. An improvement in this direction involves performing SA at a topic level (i.e., identifying the different topics of a text/source and attributing a sentiment score to each one), but this technique still remains a challenge.

Although polarity is the most widely accepted approach in SA, it might be too limited for teaching and learning. Complementary, from the 22 studies considered, four studies identify sentiments or emotions from existing classifications, mostly as a complement to polarity. Sentiment identification is based on different psychological models. For example, Alencar and Netto (2020) and Nunez (2020) identify emotion based on Plutchik's (1984) model (anger, fear, sadness, disgust, surprise, anticipation, trust, and joy), Featherstone and Botha (2015) uses Ekman's (1992) 6 affect categories (Joy, surprise, fear, sad, anger, disgust), while Osorio Angel et al. (2020) characterizes a set of emotions arising in learning with six axes, based on Kort et al. (2001). Integrating those models into SA can contribute to evidence the relationship between a significant number of emotional states and learning outcomes. As a matter of fact, SA would improve in accuracy if emotions were considered.

Finally, in some studies, as in Yu et al. (2018), human raters are involved. While this can be a good practice to ensure accuracy and validity of the SA method, to this purpose we also suggest considering involving some of the own subjects in future controlled experiments. In formal settings as in higher education platforms, the scarcity of students' messages or the communication methods can involve significant limitations.

RQ2. Frontend software based on SA developed

Only six out of the 22 studies have designed a specific software for visually representing SA results. From these, two different types of visual tools emerge. The first is designed for researchers use (two studies), as in Alencar and Netto (2020) and Nunez (2020). These representations are based on group information, involve the use of complex bar, pie, and sentiments' graphs, and are not timely (perform the analysis when the activity has ended).

The second type of visual tools (four studies) is designed and implemented for final users (teachers or students) and provides timely information with complementary analysis. For example, through the SA architecture described in Cobos et al. (2019) (Fig. 3) these authors developed a tool to provide visual information of SA at a learners' individual level (Fig. 4). Their tool allows to download the results. Elia et al. (2019) designed RAMS (Rapid monitoring of learners' satisfaction), a visual analytics software (pie charts with polarity trends) for LA and SA, with a dashboard for course and individual students' levels, that also performs cluster analysis. Wang and Zhang (2020) present through a Network and text representations, students' sentiment tendencies. These tools perform timely analysis and, although they offer complementary functionalities such as cluster analysis or downloadable outputs, the

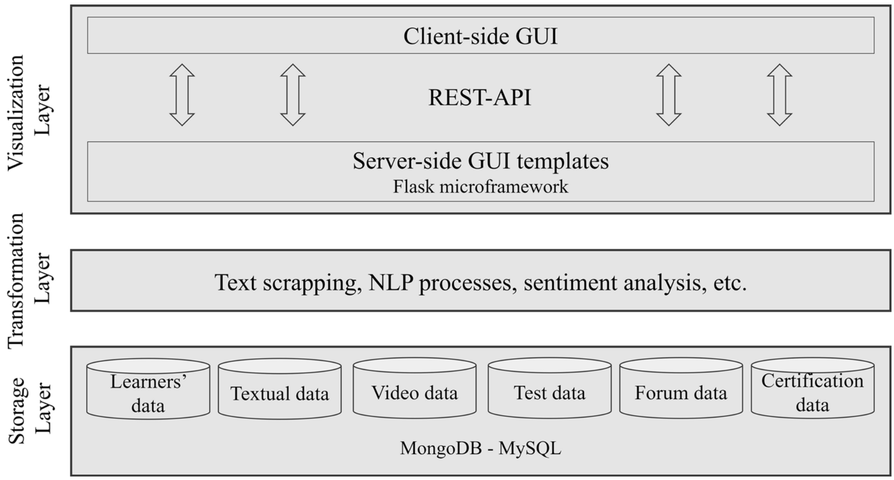

Fig. 3 Architecture of EdX-CAS for SA, adapted from Cobos et al. (2019)

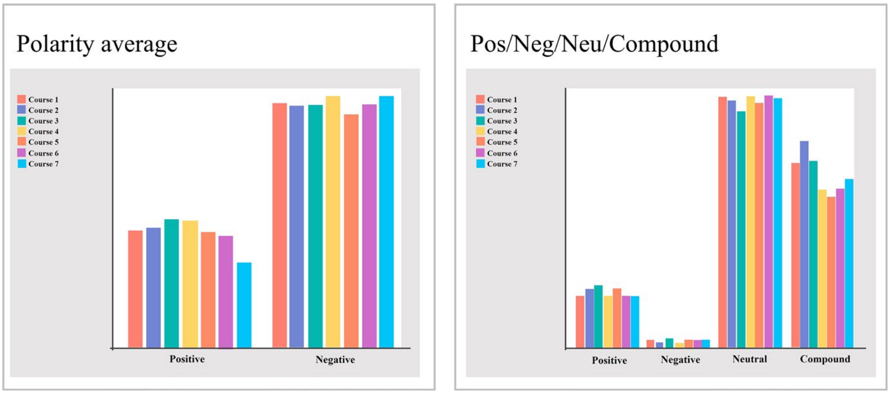

Fig. 4 Example of the results of the visual tool EdX-CAS showing the polarity of a learner's opinions in different courses, adapted from Cobos et al. (2019). *Note.* Teachers can click on any course id (at the left side of each chart) to display the particular polarity results

amount of information presented is still big and can be overwhelming for teachers. We believe that future frontend solutions might need to display information clearly and in an understandable way, allow teachers to process data in real-time, and show trends in data over time. Based on these features, the proposal made by Yu et al. (2018), presenting a dynamic diagnostic and self-regulated system in a speedometer visual dashboard combining SA and LA at a student level, can constitute a good example of frontend applications for educational purposes (see Fig. 5).

Hence, although only four studies present frontend solution, many other studies analyzed in this research acknowledge that the next steps in SA should go towards the development of more easy-to-use visual analytics tools for teachers and students, both at individual and at a social and collaborative levels, that would help instructors providing effective feedback interventions (Dehbozorgi et al., 2020; Le et al., 2018). This development would benefit from integrating usability tests for teachers to maximize SA tools for personalizing the information displayed by the user, such as dynamic filters (by students' group, by each student, by academic year…), dynamic grouping (so teachers can group their own charts and add notes) and dynamic data labels (so teachers can more easily read the data).

RQ3. General contributions of SA as a tool for students' assessment in higher education

Three main contributions of SA are discussed, based on the findings:

Assessing the emotional climate of students about an educational intervention

SA techniques are mostly used to assess the emotional climate of students in higher education, as in Spatiotis et al. (2018), or related to one particular topic, such as the use of Mobile phones in learning, as in Abdulsalami et al. (2017). Following this approach, in some studies the emotional climate assessment is carried out to

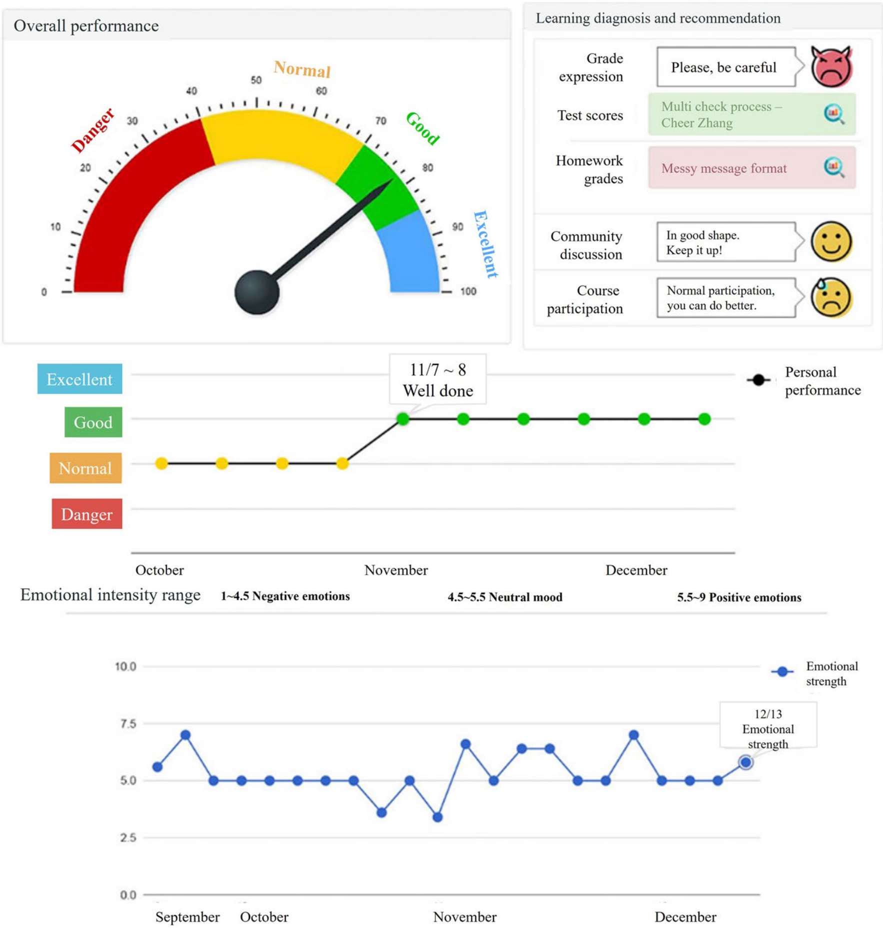

Fig. 5 Dynamic diagnostic and self-regulated (DDS) system (dashboard on top left; diagnostic and suggestion report on top right; weekly status on the middle; emotional valence value on the bottom) translated from Yu et al. (2018)

evaluate the impact of an educational intervention, such as the use of gamification strategies in higher education in Bilro et al. (2022), Featherstone and Botha (2015) and Mostafa (2020), the implementation of practical and hands-on activities in Suwal and Singh (2018), or the comparison between online or hybrid learning in Camacho and Goel (2018). These studies evidence that, although SA needs to overcome important limitations mentioned in previous sub-sections, SA can help higher education teachers and researchers to have quick results of the implementation of educational innovations complementary to traditional students' gradings, releasing research teams from the burden of analyzing a big amount of data in a first screening phase. Moreover, as SA measurements can be easily performed periodically, the evolution of the emotional climate or state can be studied to identify behavioral patterns, as in Chaabi et al. (2019) and Osorio Angel et al. (2020).

Results also corroborate that the measurement of the emotional climate through SA techniques can be a useful tool to identify gender biases. Hence, Abdulsalami et al. (2017) and Nunez (2020) described that male students expressed more positive comments than their female peers in VLE, while Okoye et al. (2020) found that students evaluated their teachers differently based on teachers' gender. These results evidence how SA can contribute to the promotion of more equitable assessment practices in higher education institutions (not only related to gender, but to any personal factor of discrimination such as race). However, this application is still underdeveloped as only three articles included gender as a variable in their study.

Improving the prediction of students' learning performance

Complementary, in four studies SA has been used to predict learning performance and possible withdrawals or dropouts early detection. Dehbozorgi et al. (2020) established a relationship between students' performance and a positive emotional climate. However, no correlation between students' negative sentiments and individual performance was measured, suggesting a need to develop more sophisticated predictive models. To this end, Alblawi and Alhamed (2017), Gkontzis et al. (2020), and Yu et al. (2018) developed predictive models combining SA of unstructured data such as students' comments or productions with structured data such as attendance, homework completion, previous grades… etc. as the one displayed in Fig. 6 made by Gkontzis et al. (2020).

These three studies show that the absence of unstructured or structured data weakens the predictive ability of the models. Particularly, Yu et al. (2018) discuss that the use of SA significantly improves this predictive ability for earlier stages over the duration of a course of instruction, as in these stages, structured data might be relatively too limited to provide enough information for prediction. Moreover, Gkontzis et al. (2020) advocate for the inclusion of teachers' data (e.g. number of forum interactions, SA…) to increase the effectiveness of predictive models. Therefore, SA predictive models need to be built from complex algorithms processing different types of data, which imply a challenging development task that many higher education institutions might not be able to assume. Finally, ethical considerations

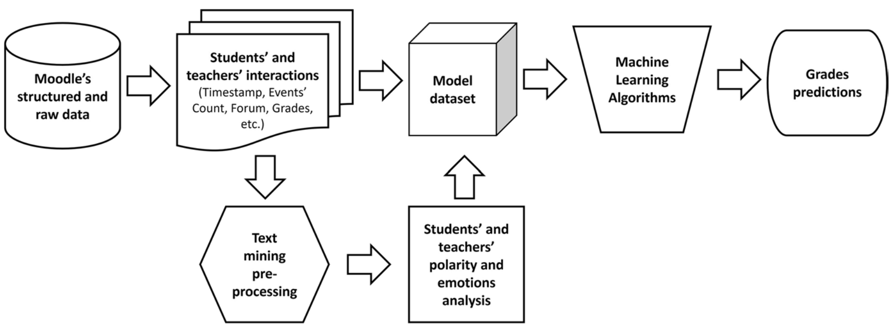

Fig. 6 Process of SA for predicting learning outcomes based on the work of Gkontzis et al. (2020)

about the use of students' data would also need to be considered, following some of the reflections shared above.

Enhancing teaching methods and feedback

Finally, SA was specifically used in two studies to improve teaching methods and feedback. Hence, Le et al. (2018) developed a module to notify teachers when a student was experiencing difficulties, as they argued that negative, distressed, and questioning sentiments can determine when a student is experiencing those difficulties. Complementary, SA of teachers' productions (e.g., videos or pdf documents) developed by Cobos et al. (2019) could help them make informed decisions whether or not it would be useful to modify their educational materials. Drawing from previous consensus in the literature, both feedback and assessment are equally important in the development of students' competencies (Pinger et al., 2018) and, therefore, it might be advisable to implement both approaches while using SA. In this sense, future studies might consider assessing the effect of SA techniques on teachers' methods to better understand which type of SA information and how is presented can better assist teachers.

Conclusions

The aim of this SLR was to understand the state of the art on SA online and hybrid learning environments in higher education related to assessment. Findings from this review show that there is a growing field of research on SA. Most of the papers are written from a technical perspective and published in journals related to digital technologies. This research focuses on teachers and institutional evaluation, especially targeting online learners' assessment and collaborative environments.

Reviewed studies mainly assess the polarity of sentiments and emotions, providing information about students' emotional climate or state. Two major approaches to SA emerge, regarding to techniques: NLP and ML. While the ML methods report higher classification accuracy, NLP is the most used due to the amount of open and free linguistic online libraries available, in despite of the significant limitations in highly specialized communication channels such as VLE in higher education. Results show that a hybrid approach is more appropriate because it grants an accurate analysis from each language and in particular on learning contexts, overcoming method limitations in isolation. However, research conducted in other languages and with more data will be useful to make a stronger corpus and more precise predictions, as well as deepen into the measurement of sentiments following different models, such as Kort's et al. (2001). Moreover, automatic data collection and preprocessing might be one of the targets to ensure appropriate integration into VLE.

Further and most relevant steps on this field are aligned towards the development of visual tools for teachers and learners of SA results. At this moment, there are few studies which have already developed and implemented frontend software in higher education contexts, showing processed results in a more visual and understandable manner, but there is still room for improvement. These tools evidence the development of new feedback systems that could possibly have better application in tracing and tracking students learning needs over time.

Summing up, major contributions of SA in higher education assessment are directed towards assessing the emotional climate or status in relation to an educational intervention, so teachers can have timely information to improve students' performance (Dehbozorgi et al., 2020; Md Faridee & Janeja, 2019), predict students' grades (Alblawi & Alhamed, 2017), and dispose of early signs to identify at-risk students (Zhang et al., 2012). We argue that including a gender perspective in developing SA techniques should contribute to the promotion of more equitable assessment. Although studies evidence different gender patterns in how students express themselves, this is an aspect that remains understudied.

Limitations on this study relate to SLR methodology such as time length and indexed databases search. Applying this methodology may have omitted other sources such as technical reports, as long as SA is a recent topic used in learning contexts. This study makes a significant contribution to research, providing practical information to higher education teachers and administrators related to online and hybrid learning students' assessment. More specifically, it calls for the urgent development of an easy-to-use, timely SA frontend tools that help teachers improving students' feedback and awareness of their sentiments in the VLE in highly specialized and more equitable learning contexts, which might result into higher learning performance.

Appendix 1: Search strategies

Scopus.

(TITLE-ABS-KEY("distance education") OR TITLE-ABS-KEY("distance learning") OR TITLE-ABS-KEY("virtual university") OR TITLE-ABS-KEY("virtual classroom") OR TITLE-ABS-KEY("online learning") OR TITLE-ABS-KEY("Web-based learning") OR TITLE-ABS-KEY("electronic learning") OR TITLE-ABS-KEY(blended) OR TITLE-ABS-KEY(hybrid) OR TITLE-ABS-KEY("computer-assisted instruction")) AND (TITLE-ABS-KEY("sentiment*") OR TITLE-ABS-KEY("sentiment analysis") OR TITLE-ABS-KEY("learning analytic*")) AND (TITLE-ABS-KEY("assessment") OR TITLE-ABS-KEY("feedback*") OR TITLE-ABS-KEY("evaluation*") OR TITLE-ABS-KEY("formative assessment")).

Wos.

(TITLE-ABS-KEY ("distance education") OR TITLE-ABS-KEY ("distance learning") OR TITLE-ABS-KEY ("virtual university") OR TITLE-ABS-KEY ("virtual classroom") OR TITLE-ABS-KEY ("online learning") OR TITLE-ABS-KEY ("Web-based learning") OR TITLE-ABS-KEY ("electronic learning") OR TITLE-ABS-KEY (blended) OR TITLE-ABS-KEY (hybrid) OR TITLE-ABS-KEY ("computer-assisted instruction")) AND (TITLE-ABS-KEY ("sentiment*") OR

TITLE-ABS-KEY ("sentiment analysis") OR TITLE-ABS-KEY ("learning analytic*")) AND (TITLE-ABS-KEY ("assessment") OR TITLE-ABS-KEY ("feedback*") OR TITLE-ABS-KEY ("evaluation*") OR TITLE-ABS-KEY ("formative assessment")) AND (TITLE-ABS-KEY ("higher education") OR TITLE-ABS-KEY ("university") OR TITLE-ABS-KEY ("adult learn*")) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "ch")) AND (EXCLUDE (LANGUAGE, "German") OR EXCLUDE (LANGUAGE, "Russian")).

109 results.

Eric – Accessed through EBSCOhost.

AB (("distance education" OR "distance learning" OR "virtual learning” OR "virtual classroom" OR "online learning" OR "Web-based learning" OR "electronic learning" OR “elearn*” OR “blended” OR “hybrid” OR "computer-assisted instruction")) AND AB ((“sentiment*” OR "sentiment analysis" OR “SA” OR "learning analytics*")) AND AB (assessment OR feedback* OR evaluation* OR "formative assessment").

23 refs. From 2006 accessible and peer review.

Gender.

WoS.

You searched for: TOPIC: (("distance education" OR "distance learning" OR "open learning" OR "virtual university" OR "virtual classroom" OR "online learning" OR "Web-based learning" OR "electronic learning" OR "elearn*" OR "blended" OR "hybrid" OR "computer-assisted instruction") AND ("sentiment*" OR "sentiment analysis" OR "learning analytics*" OR "analytics*" OR "teaching analytics*" OR "big data") AND ("assessment" OR "feedback*" OR "evaluation*" OR "formative assessment" OR "formative*" OR "immediate*" OR "achieve*")) AND (“gender” OR “women” OR “female” OR “girl*”).

Refined by: RESEARCH DOMAINS: (SOCIAL SCIENCES OR ARTS HUMANITIES) AND **DOCUMENT TYPES:** (ARTICLE OR MEETING OR REVIEW) AND **RESEARCH AREAS:** (EDUCATION EDUCATIONAL RESEARCH OR COMPUTER SCIENCE OR COMMUNICATION OR PSYCHOLOGY OR BEHAVIORAL SCIENCES OR SOCIAL SCIENCES OTHER TOPICS).

Timespan: 2010–2020. **Databases:** WOS, CCC, DIIDW, KJD, MEDLINE, RSCI, SCIELO.

Search language = Auto.

Total: 16 refs.

AND (TITLE-ABS-KEY (*"gender"*) OR TITLE-ABS-KEY (*"wom?n"*) OR TITLE-ABS-KEY (*"girl*"*) OR TITLE-ABS-KEY (*"female*"*)).

12 results.

Eric—Accessed through EBSCOhost.

AND AB (assessment OR feedback* OR evaluation* OR "formative assessment") AND AB ((gender OR women OR female* OR girl*)).

Appendix 2: Summary of findings

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
1. Abdulsalami et al. (2017) C	Experimental, Cross-sectional	Blended. 358 (2nd, 3rd and 4th year) undergraduate computer science students in physical science (Nigeria)	SA on students' perception on the use of smartphones and its relationship with academic performance. Relationship between sentiment polarity and the academic level of students	Binary polarity: positive / negative	NLP Sentence level	Bar diagrams, not timely, polarity score related to 3 outcome variables	Are emotions in students written text related to gender?
2. Alblawi and Alhamed (2017) C	Theoretical and technical solution	Blended. Pilot implementation in a programming course HE. 500 students' 1st semester marks and opinions of the programming course. (Saudi Arabia)	1) Academic Integration (i.e. student GPA; marks; satisfaction with academic experience; and interest in classes and programs). 2) Social Integration (i.e. relationship with other students; impact of peer group…). 3) Institutional Commitment (i.e. funding; physical facilities…). 4) Out-of-institution factors (i.e. finance; health…)	Polarity 3-level scale: positive, negative, neutral	Hybrid: NLP and ML Sentence level	Algorithm presenting sentiment score as the polarity (negative, positive, neutral) for each student. Not timely. Improves 10% the model only based on Structured Data FUTURE: produce an integrated LA solution for making decisions concerning student retention rates and methods to improve their performance	No

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
3. Alencar and Netto (2020) C	Experimental	Online (VLE) 160 Students of the technical specialization course in Information Technology Management, in the discipline "Computer Network Management", (Brazil)	Performance of students in tasks and activities proposed by the teacher: Forum and Assignment	Binary polarity: positive / negative Emotion: admiration, joy, interest, anger, disgust, sadness, surprise and fear	Hybrid: NLP and ML Sentence level	A system integrated with the VLE using a multi-agent system to collect messages posted by students in discussion forums and through SA to identify the Emotional state of the class and present it as pie charts, not timely	NO
4. Bilro et al. (2021) J	Qualitative, Quasi-experimental	Blended Gamification in a semester, 91 undergraduate students (Portugal)	The role of gamification in Students' engagement. Cognitive engagement as the attention and effort of students to interact with other students and tutors via discussion forums, comments, and posts	Polarity 5-level scale: (P+, P, Neutral, N, N+)	NLP Message level	Rate of negative and positive polarity of the activity. Not timely, not graphical	No (only the % of female students is mentioned)

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
5. Camacho and Goel (2018) C	Longitudinal study	Online 2014–2017 courses (71,671 posts) Leverage SA and readability assessments to quantify the evolving quality of discourse on the online forum discussions of the various sections. (USA)	Learning as part is a social process: communication and collaboration are critical to learning. Collaborative learning leads to an increased sense of community, skill acquisition, and better learning outcomes in an online environment. Online class discussion forums	Polarity 3-level scale: positive, negative, neutral Additional normalized compound: -1 (extreme negativity) to + 1 (extreme positivity)	NLP Sentence level	Rate of negative and positive polarity of the activity. Not timely, not graphical Longitudinal analysis	No
6. Chaabi, Khadija and Mounia, (2019) J	Theoretical and technical solution	Online 120 messages of Master Students Technology (Morocco)	CSCL Asynchronous activities (emails, forums)	Behavioral profiles: Animator, Checker, Seeker, Independent	ML Word level	Classification of behavioral profiles (Animator, checker, seeker and independent) of learners in social media discussion groups	NO

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
7. Cobos, Jurado and Blázquez-Herranz (2019) J	Technical solution	Online university courses, 7 MOOCs. (Spain)	Analyze the contents of online courses (video transcriptions, readings, questions and answers of the evaluation activities and learner's posts in forums…) and the contributions of learners to improve teaching material and teaching–learning processes	Polarity 3-level scale: positive, negative, neutral Level of objectivity	NLP Sentence level	Design and development of *edX-CAS*. "Content Analyser System for edX MOOCs". Visual (bar diagrams) display and download at i) Learners level, ii) Forum level "Social" button and iii) other texts Data analyses by clicking the "General" button	NO
8. Dehbozorgi et al. (2020) C	Experimental	Blended Introduction to Computer science 28 students (USA)	Collaborative learning A positive correlation between students' performance and their positive sentiment as well as the level of subjectivity in speech (verbal conversations emotional state)	Polarity 3-level scale: positive, negative, neutral Normalized score: 0 (extreme negativity) to + 1 (extreme positivity)	NLP Sentence level	Future: serve as a performance predictor in earlier stages of the semester to provide timely feedback to students and enables instructors to make interventions that can lead to student success	NO

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
9. Elia et al. (2019) J	Design science research	Online Collaborative Learning Programming students (Italy)	CL: Analyze individual students' contributions in online learning activities and assess the level of satisfaction towards the course. Real-time monitoring of LS Real time monitoring of LS may support course delivery and reduce early dropouts	Binary polarity: positive / negative	ML Word level	*RAMS* Software with visual analytics (pie charts with polarity) component for LA and SA in real-time individual contributions to online CL	NO
10. Md Faridee and Janeja (2019) C	Experimental	Blended 300 IS students (USA)	Students receiving peer mentoring can express more complex ideas with fewer words and thereby receive higher grades by the end of the semester (grades of exercises to help increase interest in data analytics in cybersecurity)	Polarity scale between -5 and 5 Scores between -1, 1 assumed to be neutral Level of objectivity	NLP Word level	No software but results in bars to compare between control and experimental group, not timely, not individual	Lack of diversity, gender and ethnicity analysis of results. FUTURE studies measuring gender

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
11. Featherstone and Botha (2015) C	Qualitative	Blended Use of tablets, gamification 136 Teacher education students (South Africa)	Badge evaluator evidence provided and awarded the badge or gave meaningful input on improvements. If all the modules presented, the teacher graduates and earns a tablet, should they have achieved the minimum criteria	Binary polarity: positive / negative Ekman's 6 affect categories (Joy, surprise, fear, sad, anger, disgust) Valence of affect (positive / negative)	NLP Word level	Bar diagrams for modules and affect categories, not for students, not timely	NO
12. Gkontzis et al. (2020). J	Experimental	Online LMS (Moodle) 30 tutors, 460 students, 6 periods of postgraduate course (Greece)	Emotions and polarity of participants in distance learning help predicting students' tendencies to abandon a lesson and in developing a PLE. Important influence of written assignments Tutor loyalty and student maturity in the learning process have a significant impact in learning	Polarity 3-level scale: negative, neutral, positive	Hybrid: NPL & ML Sentence level	Implementation of DM methodologies to provide timely, personalized and accurately the scores of eight classes, not software or front-end tool yet Sentiments variables	NO

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
13. Le et al. (2018) C	Big data analysis EPDM	Blended 1st year computer science students 4666 sentences (Australia)	Discussion forums in university courses to support students. Research has begun to explore Text classification to forum text to determine if lecturers can be notified (identifying negative, distressed, and questioning sentiments)	Problematic post: non-critical/ critical	ML Message level	A classification tool that identifies problematic posts into non-critical, when there is no problem expressed by the student, and critical, when there was evidence of an issue associated with learning Not timeless FUTURE: highly desirable to develop tools that educators can efficiently and easily use to explore online student data	NO
14. Mostafa (2020) C	Quantiative	Online Gamification 1000 students in Science & Technology transport university (Egypt)	Student motivation and learning engagement of students in HE	Agreement / Disagreement sentiments	Hybrid: NLP & ML Message level	SA classifier to analyze students sentiments while using gamification tool. No representation, not timely	NO

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
15. Nunez (2020) J	Quantitative (SA) EPDM	Blended 37 students Agriculture Course (USA)	Hands-on activities in a LMS: After each activity, students submitted a 200-word reflective essay about their experience with it, to evaluate hands-on activities and identify those that promote student learning and those that need further refinement	Binary polarity: positive / negative Emotional states: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust (Plutchik, 1994)	NLP Word level	Bar, pie, and other graphs for teachers, not timely, group level	Present sample demographics gender: Gender only affected overall sentiment in activity 5, where male students expressed more positive sentiments than female students

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
16. Okoye et al. (2020) J	Quantitative EPDM	Online n1 = 52,218 comments for male teachers and n2 = 41,072 for female teachers (México)	Educational Process and Data Mining (EPDM) model: leverage the opinions of students to give useful feedback that can be used to enhance the end-to-end processes within the educational domain It is essential to monitor the students' feelings or emotions because learners with a positive attitude have been seen as more confident and motivated to learn (Weston et al., 2019)	Polarity 3-level scale: positive, negative, neutral Analysis of emotional valence (-4,0,8)	NLP Word and sentence level	Word clouds & sentiment Bar charts. Real time EPDM implementation, group level	Identify terms used by the student to describe the teachers' qualities at the end of the semester. The analysis consider gender differences in teachers' attributes SA to determine the intensity of the students' comments towards the teachers by gender Implications of the statistical significance and differences by the gender of the teachers

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
17. Osorio, Negrón & Valdez (2020) C	Exploratory study. (Interpretative descriptive analysis)	Online CMC in VLE (Moodle) 56 students of engineering, 361 forum messages. (México)	Affective Computing aim is to create computer environments for learning capable of promoting positive emotional experiences that improve student's performance in a Kort model for STEM students in HE, discussion forums (reflective and cognitive intentions)	6×6 axes of emotion from the students in technology by Kort (2001)	Hybrid: NLP and ML Sentence level	A conceptual model for the analysis of emotions in Spanish, intended for Engineering students during their learning process in a collaborative computer environment FUTURE: developing the computational model that automates this Process	NO

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
18. Spatiotis et al. (2018) C	Big data analysis E(P)DM	Blended 2600 teachers 11,156 opinions (Greece)	Determine how students learn and identify how the learning process could contribute to the performance of students EDM: discover hidden knowledge, understanding students' needs and methods to improve the students' experience, activities and performance. Interesting activities can be proposed for students related to their interactions with the online learning tools, thus contributing to their high performance	Polarity 5-level scale, from 5: very positive to 1: very negative	ML Message level	An opinion mining platform which is able to classify participants' opinions in education lessons according to their polarity and to analyze them in order to contribute to improvement of the teaching procedure	NO

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
19. Suwal and Singh (2018) J	Mixed methods	Online 65 students in a post-graduate Engineering course (Finland)	Students' feedback and perceptions are critical to make an informed decision about the changes: desirable to obtain feedback and analyze it methodologically. SA to assess student feedback through written assignments	Polarity 3-level scale: positive, negative, neutral	NLP Sentence level	Integrated feedback systems have the possibility of using automated SA has potential for active development of new feedback systems that could have better use of tracing and tracking the development needs of courses over time	NO
20. Wang and Zhang (2020) J	Theoretical and technical solution	Online Learning community 171,430 comments (China)	SA can reflect the students' learning situation, thus provide the theoretical basis for following revision of teaching plans	Polarity 3-level scale: positive, negative, neutral	Hybrid: NLP and ML Topic level	Network and text representations based on open interactive visualization library pyLDAvis and the open-source tool Colibri/MLA A new effective model for topic SA perceives students' sentiment tendencies and provides powerful practical reference for teachers	NO

Ref (APA)	Study	Context (online/face to face / blended)	Learning assessment	Polarity / Sentiment / Emotion measurement	SA technique and granularity	SA visualization tool	Gender analysis
21. Yu et al. (2018) J	Theoretical and technical solution	Blended 181 undergraduate students Introduction to Computer Science (Taiwan)	improvement in teaching and learning is important to maintain students' performances and the effectiveness of the learning process. SA to identify affective information from text-based self-evaluated comments written by students (containing rich emotional information)	Polarity 3-level scale: positive, negative, neutral	Hybrid: NLP and ML Message level	Speedometer Dynamic diagnostic and self-regulated system Timely and student level weekly status, emotional valence value Provide real-time feedback to educators, to help students become self-regulated learners	NO
22. Zhang et al. (2012) J	Mixed methods	Blended Medical students (postgraduate) 812 statements (USA)	Explore SA to assess the feasibility and value of an automated approach for synthesizing evaluation comments, to make better assessments of students at a ACGME competency level	Polarity 3-level scale: Extremely positive, positive, and non-positive	ML Sentence level	FUTURE: the potential to provide medical educators researchers with automated tools to synthesize qualitative evaluation data in a streamlined and efficient manner	NO

Acknowledgements We would like to thank Yonathan Borden who provided comments on the early draft of this paper. We would also like to thank the reviewers of this article who, through their insightful advice, they helped us to improve our work.

Funding Open access funding provided by Universitat Rovira i Virgili. This research has been conducted in the context of "MindGAP". [LCF/PR/SR19/52540001] and within the research group ARGET research group [2021SGR00707].

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

References

*References included in the systematic literature review

*Abdulsalami, A. O., Ahmad, B. I., Umar, M. A., Abubakar, A. H., Jauro, F., Kufena, A. M., & Ekoja, E. A. (2017). Sentiment analysis of students' perception on the use of smartphones: A cross sectional study. In *2017 second international conference on informatics and computing (ICIC)*, *2018-Janua*, 1–5. https://doi.org/10.1109/IAC.2017.8280625

*Alblawi, A. S., & Alhamed, A. A. (2017). Big data and learning analytics in higher education. In B. Kei Daniel (Ed.), *Big data and learning analytics in higher education*. Springer International Publishing. https://doi.org/10.1007/978-3-319-06520-5

*Alencar M, Netto J (2020) Measuring student emotions in an online learning environment. In: *Proceedings of the 12th international conference on agents and artificial intelligence*, vol 2, pp 563–569. https://doi.org/10.5220/0008956505630569

Andreotta, A. J., Kirkham, N., & Rizzi, M. (2021). AI, big data, and the future of consent. *AI and Society*. https://doi.org/10.1007/s00146-021-01262-5

*Bilro, R. G., Loureiro, S. M. C., & Angelino, F. J. de A. (2022). The role of creative communications and gamification in student engagement in higher education: A sentiment analysis approach. *Journal of Creative Communications*, *17*(1), 7–21. https://doi.org/10.1177/0973258621992644

Borokhovski, E., Tamim, R., Bernard, R. M., Abrami, P. C., & Sokolovskaya, A. (2012). Are contextual and designed student–student interaction treatments equally effective in distance education? *Distance Education, 33*(3), 311–329. https://doi.org/10.1080/01587919.2012.723162

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society, 15*(3), 3–26.

*Camacho, I., & Goel, A. (2018). Longitudinal trends in sentiment polarity and readability of an online masters of computer science course. *Proceedings of the 5th annual ACM conference on learning at scale, L at S 2018*, *January 2014*, 0–3. https://doi.org/10.1145/3231644.3231679

*Chaabi, Y., Lekdioui, K., & Boumediane, M. (2019). Semantic analysis of conversations and fuzzy logic for the identification of behavioral profiles on facebook social network. *International Journal of Emerging Technologies in Learning (IJET)*, *14*(07), 144. https://doi.org/10.3991/ijet.v14i07.8832

Chiarello, F., Bonaccorsi, A., & Fantoni, G. (2020). Technical sentiment analysis. Measuring advantages and drawbacks of new products using social media. *Computers in Industry, 123*, 103299. https://doi.org/10.1016/j.compind.2020.103299

Çoban, Ö., Özel, S. A., & Inan, A. (2021). Deep learning-based sentiment analysis of facebook data: The case of Turkish users. *The Computer Journal, 64*(3), 473–499. https://doi.org/10.1093/comjnl/bxaa172

*Cobos, R., Jurado, F., & Blazquez-Herranz, A. (2019). A content analysis system that supports sentiment analysis for subjectivity and polarity detection in online courses. *IEEE Revista Iberoamericana de Tecnologias Del Aprendizaje*, *14*(4), 177–187. https://doi.org/10.1109/RITA.2019.2952298

*Dehbozorgi, N., Lou Maher, M., & Dorodchi, M. (2020). Sentiment analysis on conversations in collaborative active learning as an early predictor of performance. In: *2020 IEEE frontiers in education conference (FIE)*, pp. 1–9. https://doi.org/10.1109/FIE44824.2020.9274119

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3–4), 169–200. https://doi.org/10.2307/j.ctv21d641d.14

*Elia, G., Solazzo, G., Lorenzo, G., & Passiante, G. (2019). Assessing learners' satisfaction in collaborative online courses through a big data approach. *Computers in Human Behavior*, *92*, 589–599. https://doi.org/10.1016/j.chb.2018.04.033

*Featherstone, C., & Botha, A. (2015). Sentiment analysis of the ICT4Rural Education teacher professional development course. In *2015 IST-Africa conference*, *May*, 1–12. https://doi.org/10.1109/ISTAFRICA.2015.7190592

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004

*Gkontzis, A. F., Kotsiantis, S., Kalles, Di., Panagiotakopoulos, C. T., & Verykios, V. S. (2020). Polarity, emotions and online activity of students and tutors as features in predicting grades. *Intelligent Decision Technologies*, *14*(3), 409–436. https://doi.org/10.3233/IDT-190137

Iglesias i Estradé, A. M. (2019). Relació entre l'anàlisi del sentiment, els resultats acadèmics i l'abandonament als entorns virtuals d'aprenentatge. *Universitas Tarraconensis. Revista de Ciències de l'Educació*, *1*(1), 19. https://doi.org/10.17345/ute.2019.1.2622

Kashy-Rosenbaum, G., Kaplan, O., & Israel-Cohen, Y. (2018). Predicting academic achievement by class-level emotions and perceived homeroom teachers' emotional support. *Psychol Schools, 55*(7), 770–782. https://doi.org/10.1002/pits.22140

Kastrati, Z., Dalipi, F., Imran, A. S., Pireva Nuci, K., & Wani, M. A. (2021). Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study. *Applied Sciences, 11*(9), 3986. https://doi.org/10.3390/app11093986

Kickmeier-Rust, M. D., Hillemann, E. C., & Albert, D. (2014). Gamification and smart feedback: Experiences with a primary school level math app. *International Journal of Game-Based Learning, 4*(3), 35–46. https://doi.org/10.4018/ijgbl.2014070104

Kitchenham, B. (2004). Procedures for performing systematic reviews. Joint technical report. http://www.inf.ufsc.br/aldo.vw/kitchenham.pdf

Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. In *Proceedings IEEE international conference on advanced learning technologies*, 43–46. https://doi.org/10.1109/ICALT.2001.943850

*Le, K., Tarmazdi, H., Vivian, R., Falkner, K., Szabo, C., & Falkner, N. (2018). Directing teacher focus in computer science online learning environments. In *2018 International conference on learning and teaching in computing and engineering (LaTICE)*, *8*(12), 119–125. https://doi.org/10.1109/LaTICE.2018.00014

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *339*, b2700–b2700. https://doi.org/10.1136/bmj.b2700

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, pp. 31–40

Martínez Cámara, E., García Cumbreras, M. Á., Villena Román, J., & García Morera, J. (2016). TASS 2015—The evolution of the Spanish opinion mining systems. *Procesamiento De Lenguaje Natural, 56*, 33–40.

McCarthy, J. (2017). Enhancing feedback in higher education: Students' attitudes towards online and in-class formative assessment feedback models. *Active Learning in Higher Education, 18*(2), 127–141. https://doi.org/10.1177/1469787417707615

McLaughlin, T., & Yan, Z. (2017). Diverse delivery methods and strong psychological benefits: A review of online formative assessment. *Journal of Computer Assisted Learning, 33*(6), 562–574. https://doi.org/10.1111/jcal.12200

*Md Faridee, A. Z., & Janeja, V. (2019). Measuring peer mentoring effectiveness in computing courses: A case study in data analytics for cybersecurity. In *2019 26th international conference on high performance computing, data and analytics workshop (HiPCW)*, 71–77. https://doi.org/10.1109/HiPCW.2019.00024

Mite-Baidal, K., Delgado-Vera, C., Solís-Avilés, E., Espinoza, A. H., Ortiz-Zambrano, J., & Varela-Tapia, E. (2018). Sentiment analysis in education domain: A systematic literature review. *Communications in Computer and Information Science, 883*, 285–297. https://doi.org/10.1007/978-3-030-00940-3_21

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 339*(7716), 332–336. https://doi.org/10.1136/bmj.b2535

Moreno-Marcos, P. M., Alario-Hoyos, C., Munoz-Merino, P. J., Estevez-Ayres, I., & Kloos, C. D. (2019). A learning analytics methodology for understanding social interactions in MOOCs. *IEEE Transactions on Learning Technologies, 12*(4), 442–455. https://doi.org/10.1109/TLT.2018.2883419

*Mostafa, L. (2020). Student sentiment analysis using gamification for education context. In *Advances in intelligent systems and computing* (Vol. 1058, pp. 329–339). Springer International Publishing. https://doi.org/10.1007/978-3-030-31129-2_30

*Nunez, G. H. (2020). Using text mining to gauge student sentiments about hands-on activities in a protected agriculture course. *HortTechnology*, *30*(1), 129–134. https://doi.org/10.21273/HORTTECH04527-19

*Okoye, K., Arrona-Palacios, A., Camacho-Zuñiga, C., Hammout, N., Nakamura, E. L., Escamilla, J., & Hosseini, S. (2020). Impact of student's evaluation of teaching: a text analysis of the teachers qualities by gender. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-00224-z

Onan, A. (2021). Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. *Computer Applications in Engineering Education, 29*(3), 572–589. https://doi.org/10.1002/cae.22253

Oreski, D., & Kadoic, N. (2018). Analysis of ICT students' LMS engagement and success. In *35th International scientific conference on economic and social development*, *November*, 196–210. https://doi.org/10.4324/9780203004937

*Osorio Angel, S., Peña Pérez Negrón, A., & Espinoza Valdez, A. (2020). From a conceptual to a computational model of cognitive emotional process for engineering students. In *Advances in intelligent systems and computing* (Vol. 1071, pp. 173–186). https://doi.org/10.1007/978-3-030-33547-2_14

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Online Learning, 16*(3), 9–20. https://doi.org/10.24059/olj.v16i3.267

Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018). Interplay of formative assessment and instructional quality—interactive effects on students' mathematics achievement. *Learning Environments Research, 21*(1), 61–79. https://doi.org/10.1007/s10984-017-9240-2

Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 197–219). Psychology Press.

Shapiro, H. B., Lee, C. H., Wyman Roth, N. E., Li, K., Çetinkaya-Rundel, M., & Canelas, D. A. (2017). Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers. *Computers & Education, 110*, 35–50. https://doi.org/10.1016/j.compedu.2017.03.003

*Spatiotis, N., Perikos, I., Mporas, I., & Paraskevas, M. (2018). Evaluation of an educational training platform using text mining. In *Proceedings of the 10th hellenic conference on artificial intelligence*, pp. 1–5. https://doi.org/10.1145/3200947.3201049

*Suwal, S., & Singh, V. (2018). Assessing students' sentiments towards the use of a Building Information Modelling (BIM) learning platform in a construction project management course. *European Journal of Engineering Education*, *43*(4), 492–506. https://doi.org/10.1080/03043797.2017.1287667

Thelwall, M. (2018). Gender bias in sentiment analysis. *Online Information Review, 42*(1), 45–57. https://doi.org/10.1108/OIR-05-2017-0139

Treceñe, J. K. D. (2019). Delving the sentiments to track emotions in gender issues: A plutchik-based sentiment analysis in students' learning diaries. *International Journal of Scientific & Technology Research, 8*(12), 1134–1139. https://doi.org/10.1109/LaTICE.2018.00014

Usart, M., Grimalt-Álvaro, C., & Iglesias-Estradé, A. M. (2022). Gender-sensitive sentiment analysis for estimating the emotional climate in online teacher education. *Learning Environments Research*. https://doi.org/10.1007/s10984-022-09405-1

Van Horne, S., Curran, M., Smith, A., VanBuren, J., Zahrieh, D., Larsen, R., & Miller, R. (2018). Facilitating student success in introductory chemistry with feedback in an online platform. *Technology, Knowledge and Learning, 23*(1), 21–40. https://doi.org/10.1007/s10758-017-9341-0

Vonderwell, S. K., & Boboc, M. (2013). Promoting formative assessment in online teaching and learning. *TechTrends, 57*(4), 22–27. https://doi.org/10.1007/s11528-013-0673-x

*Wang, K., & Zhang, Y. (2020). Topic sentiment analysis in online learning community from college students. *Journal of Data and Information Science*, *5*(2), 33–61. https://doi.org/10.2478/jdis-2020-0009

Weston, T. J., Dubow, W. M., & Kaminsky, A. (2020). Predicting women's persistence in computer science- and technology-related majors from high school to college. *ACM Transactions on Computing Education, 20*(1), 1–16. https://doi.org/10.1145/3343195

Xiong, Y., & Suen, H. K. (2018). Assessment approaches in massive open online courses: Possibilities, challenges and future directions. *International Review of Education, 64*(2), 241–263. https://doi.org/10.1007/s11159-018-9710-5

Yadegaridehkordi, E., Noor, N. F. B. M., Bin-Ayub, M. N., Affal, H. B., & Hussin, N. B. (2019). Affective computing in education: a systematic review and future research. *Computers & Education, 142*, 103649. https://doi.org/10.1016/j.compedu.2019.103649

*Yu, L. C., Lee, C. W., Pan, H. I., Chou, C. Y., Chao, P. Y., Chen, Z. H., Tseng, S. F., Chan, C. L., & Lai, K. R. (2018). Improving early prediction of academic failure using sentiment analysis on self-evaluated comments. *Journal of Computer Assisted Learning*, *34*(4), 358–365. https://doi.org/10.1111/jcal.12247

*Zhang, R., Pakhomov, S., Gladding, S., Aylward, M., Borman-Shoap, E., & Melton, G. B. (2012). Automated assessment of medical training evaluation text. *AMIA ... Annual Symposium Proceedings / AMIA Symposium*, *2012*, 1459–1468

Zhou, J., & Ye, J. (2020). Sentiment analysis in education research: A review of journal publications. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2020.1826985

Carme Grimalt-Álvaro holds a PhD in Mathematics and Science Education from the Universitat Autònoma de Barcelona. She worked as a pre-doctoral and post-doctoral researcher at the Centre de Recerca per a l'Educació Científica i Matemàtica (CRECIM) in several research, dissemination, and innovation projects. She is currently a researcher at ARGET research group from Universitat Rovira i Virgili and a visiting teacher at Universitat Autònoma de Barcelona. Her work is centred on two main domains: teaching and learning with digital technologies and promoting equity in STEM education.

Mireia Usart holds a degree in Physics (Universitat de Barcelona), a MSc. In Education and ICT and a PhD in Education and ICT (Universitat Oberta de Catalunya). She is currently lecturer and researcher at the Universitat Rovira i Virgili (URV) in Tarragona (Spain). She has been awarded in international conferences and has participated and leaded different European and Spanish R&D projects since 2011. Her expertise goes from computer science (as database designer and data analysist) to psychometry in educational research, she is an expert in quantitative methods for educational research and Teacher Digital Competence and has published impact papers and book chapters in these areas.