International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Analyzing Middle School Students' Distance Education Experiences in COVID-19 via Sentiment Analysis and Topic Modeling

Ekrem Bahçekapılı[1], Bülent Kandemir[2], and Elif Baykal Kablan[3]
[1]Department of Management Information Systems, Karadeniz Technical University; [2]Department of Educational Sciences, Ordu University; [3]Department of Software Engineering, Karadeniz Technical University

Abstract

This study investigated middle school students' experiences with emergency remote education during the COVID-19 pandemic using natural language processing (NLP), sentiment analysis, and topic modeling techniques. A total of 2,739 valid responses from Turkish students (ages 9–15) were collected through open-ended survey questions regarding the perceived advantages and disadvantages of distance learning. Sentiment classification was performed using a semi-supervised machine learning approach, combining TF-IDF, Word2Vec, and FastText vectorization with five classification algorithms. The TF-IDF + support vector machines (SVM) combination yielded the highest performance (F1 = 0.85). Results show a total of 1,867 positive and 2,542 negative opinions, indicating that students generally adopted a more critical view of distance education. To explore the thematic structure of opinions, topic modeling was applied with six topics. Positive sentiments clustered around themes such as educational continuity, health protection, time savings, flexible scheduling, self-regulated learning, and digital literacy. Negative sentiments were dominated by themes including limited interaction, screen fatigue, perceived low quality, technical barriers, and structural inequalities. Findings suggest that while students appreciated the safety and flexibility of remote learning, they also faced significant pedagogical, physical, and technological challenges. The study contributes methodologically by demonstrating the effectiveness of AI-based text analysis and offers practical implications for designing more equitable and student-centered digital education models. These results underscore the importance of integrating NLP and machine learning tools into educational research to uncover deeper insights from student-generated content at scale.

Keywords: sentiment analysis, topic modeling, student emotion, student perception, COVID-19, NLP

Introduction

The COVID-19 pandemic has affected education systems worldwide, producing marked changes for K–12 learners. Reliant on structured, face-to-face classrooms and characterized by age-specific developmental needs, this cohort encountered serious challenges (Adams et al., 2024; Tomaszewski et al., 2023). The abrupt shift to remote instruction exposed critical infrastructure gaps especially in rural and socioeconomically disadvantaged regions, thereby amplifying preexisting inequalities. As the digital divide widened, meaningful participation and learning opportunities became increasingly constrained (Dhawan, 2020; Hurling et al., 2024). To interpret these effects, we first distinguish emergency remote teaching (ERT) from pedagogically designed online learning: ERT prioritizes continuity through rapid, temporary solutions under crisis constraints. This distinction frames our interpretation of students' experiences during the crisis.

Remote instruction, made compulsory by the pandemic, generated both benefits and obstacles. Digital platforms afforded flexibility and continuity in learning, showing potential for enhancing engagement and achievement (Lo et al., 2023). Yet the rapid migration online also revealed significant shortfalls in meeting individual learning needs (Diz-Otero et al., 2023). Students reported technical difficulties, loss of motivation, and a lack of collaborative environments, all of which contributed to learning deficits (Sandvik et al., 2024). At the middle school stage, these design and implementation constraints intersect with developmental needs. In particular, early adolescents are reorganizing motivation, belonging, and self-efficacy (Eccles & Roeser, 2011). Consistent with this, from a self-determination perspective, threats to autonomy, competence, and relatedness predict lower engagement in online settings (Ryan & Deci, 2000).

The pandemic likewise exerted profound social and emotional pressures on middle schoolers. Loneliness and mental health issues increased globally (Geulayov et al., 2024). Social media sustained peer communication but also spread misinformation and panic (Radwan et al., 2020). Interruptions to peer and teacher relationships elevated stress levels, especially at the middle school stage (Albert, 2024). The crisis highlighted the centrality of digital infrastructure for equitable access, especially in economically disadvantaged contexts (Hurling et al., 2024). From a Community of Inquiry (CoI) perspective, disruptions to social and teaching presence would have been expected to erode belonging and perceived support, magnifying stress and disengagement in early adolescence (Garrison, 2016; Garrison et al., 2000). Beyond connectivity, second- and third-level aspects of digital divide skills, and the translation of use into academic benefit, which varies widely across contexts, help explain uneven socioemotional outcomes under remote learning (van Dijk, 2006). These dynamics are likely to be pronounced in early adolescence.

Artificial intelligence (AI) technologies have emerged as powerful tools for making learning more effective and personalized. AI-enhanced applications analyze students' emotional states and adapt content accordingly (Lin & Chen, 2024). Facial-recognition and affect-detection systems have enabled real-time monitoring of learners' emotions (Fang et al., 2023). In addition, natural language processing (NLP) and machine learning techniques identify cognitive and affective barriers, offering timely interventions (Martínez-Comesaña et al., 2023). Personalized feedback has been shown to bolster motivation and emotional resilience (An et al., 2023), although concerns persist about excessive surveillance and equity (Lin & Chen, 2024). Aligned with the CoI framework and transactional distance, analytics-enabled, timely feedback can strengthen teaching and social presences and reduce structure–dialogue imbalances that fuel disengagement in remote contexts (Garrison, 2016; Moore, 1993). In

practice, dashboarded prompts and micro-goal check-ins can scaffold self-regulated learning processes involving planning, monitoring, and self-evaluation, particularly salient for early adolescents in flexible online settings (Pintrich, 2004; Zimmerman, 2002). Large-scale NLP analysis therefore offers an innovative alternative to traditional survey-based studies.

Against this backdrop, this study aimed to examine middle school students' perceptions of emergency remote teaching by (a) quantifying their positive versus negative sentiment and (b) extracting latent thematic structures through advanced NLP techniques, offering a scalable alternative to traditional survey-based approaches, thereby delivering theoretical, methodological, and practical contributions. As noted above, because these data were generated under ERT, we interpreted all patterns in light of crisis-driven, rapidly deployed solutions; accordingly, claims have been bounded to ERT conditions rather than fully designed online learning.

We framed our research with two questions and two sub questions:

RQ1. How are middle school students' perceptions of remote education distributed across positive and negative sentiment categories?

RQ2. What latent themes underlie middle school students' perceptions of distance education?

RQ2a. Which themes emerge from students' positive opinions?

RQ2b. Which themes emerge from students' negative opinions?

Literature Review

AI now supports learning on two fronts; it serves students directly through adaptive tutors and affect-aware dashboards, while also supplying teachers and instructional designers with indirect, data-rich insights into learners' emotions and thought processes (Lin et al., 2024; Uçar et al., 2024). NLP and machine learning pipelines analyze large corpora of student text, enabling educators to identify individual needs and tailor instructional strategies. For example, large language models (LLMs) can detect affective states in real time and summarize formative feedback, thus reducing teachers' analytical workload (Xu et al., 2024). Techniques such as sentiment analysis and text mining further feed design analytics that help instructors construct more personalized and responsive learning environments (Al Husaeni et al., 2022; Bittencourt et al., 2024). However, most prior studies rely on small-sample surveys or focus on university settings; as a result, we still know little about how middle school pupils felt during ERT. Filling that gap was the central aim of RQ1 in this study.

Recent K–12 research has shown that NLP can act both as a diagnostic lens and as an adaptive support in online or hybrid learning contexts. Using network analysis combined with topic modelling, Xing et al. (2025) found that "extra-periphery" participants in a large asynchronous mathematics forum could achieve the highest mathematical-literacy scores when interaction quality was high. At the message level, sentiment and tone analysis applied to middle school teacher feedback by Baral et al. (2023) revealed patterns that shape student motivation, while an emoji-based interface designed by Zarkadoulas and Virvou (2024) captured pupils' emotional states and highlighted gender differences in expression. From the teacher perspective, an NLP-driven virtual facilitator for professional development produced significant gains in student performance in a randomized trial reported by Copur-Gencturk et al. (2024), aligning with the "Turing Teacher" attributes outlined by Pelaez et al.

(2022). Inclusive angles are also emerging: a mixed qualitative–sentiment study by Tzimiris et al. (2023) documented layered psychological and technical barriers faced by students with functional diversity during ERT. Finally, automated text analysis developed by Žitnik and Smith (2024) flagged off-topic posts in fourth-grade book-club discussions with 90% accuracy, pointing toward real-time analytics that could keep young learners on track.

Various AI techniques are employed to analyze student opinions in remote education. One of the most prominent NLP approaches is topic modelling, which uncovers latent themes within a text. For example, Mujahid et al. (2021) used latent Dirichlet allocation (LDA) to analyse student comments on Twitter concerning online learning during the COVID-19 period. Their study revealed the main challenges of online learning, particularly infrastructure deficiencies and lack of technical support. However, LDA coherence degrades sharply with short, grammar-sparse comments, typical of K–12 data (Gallagher et al., 2017). In another study, researchers conducted a large-scale thematic analysis of social media data, allowing them to map broad perspectives on online education (Mishra et al., 2021). Waheeb et al. (2022) likewise analyzed social media datasets from the pandemic period, combining LDA with ontology-based approaches to capture both thematic and sentiment trends. Information-theoretic models such as correlation explanation (CorEx) achieve higher semantic consistency on short documents and allow researchers to anchor domain-specific keywords (Gallagher et al., 2017), yet have rarely been applied to open-ended responses from middle school students, particularly in non-English contexts.

Another major application of machine learning and NLP is sentiment analysis, which seeks to identify emotions embedded in text. Akhmedov et al. (2021) employed the joint sentiment topic (JST) model to integrate thematic and sentiment analysis, enabling a word-level examination of student opinions; this approach provided a more comprehensive understanding of perceptions related to online education. Nevertheless, JST requires large, labelled corpora resources that remain scarce for Turkish middle school datasets. Lin et al. (2024) used large databases to apply sentiment-analysis techniques that identified the key drivers of student satisfaction and negative emotions in online learning. In a separate investigation, researchers integrated machine learning and deep learning methods to classify emotions in online-course reviews with high accuracy. Such studies offer valuable insights for developing student-centered instructional strategies (Onan, 2021).

In sum, prior research has established the value of topic and sentiment modelling but left unanswered how middle school learners themselves articulate advantages and disadvantages of ERT in their own language; addressing this gap constituted the focus of RQ2, RQ2a, and RQ2b in our study.

Methodology

Convenience sampling allows researchers to obtain data quickly and easily from suitable sources. This method is particularly useful when probability sampling techniques are not feasible, enabling researchers to focus on subjects that are accessible and relevant to the study's purpose (Marshall, 1996). Although convenience sampling may introduce bias by limiting the representativeness of the sample, it was considered appropriate in this study due to the extraordinary circumstances of the COVID-19 pandemic. During this period, accessibility and timeliness were critical, and voluntary participation from available students provided a practical means to capture authentic perceptions of ERT.

The study's participants consisted of middle school students (ages 9–15, mean age = 11.95, *SD* = 1.2; a critical developmental stage and a relatively less-studied group) enrolled in the 2019–2020 academic year in Türkiye. During this period, these students were educated entirely through remote learning

methods (live classes and asynchronous learning resources) for the first time. Following ethical committee approval (E-81614018-000-330), necessary information was conveyed to the students' families through the Ministry of National Education, and a survey was distributed via an online form. Data collection was conducted anonymously. A total of 2,890 students voluntarily participated in the study. In the preliminary analysis, responses with missing data, participants who answered all questions identically, and consecutive identical responses were excluded. A total of 2,739 valid responses (female = 56.7%, male = 43.3%) were collected and accepted for further data analysis.

Data were obtained through an online form. The data collection instrument included two open-ended questions. The first question was, "What do you think are the advantages of distance education?" and the second was, "What do you think are the disadvantages of distance education?"

Data Analysis

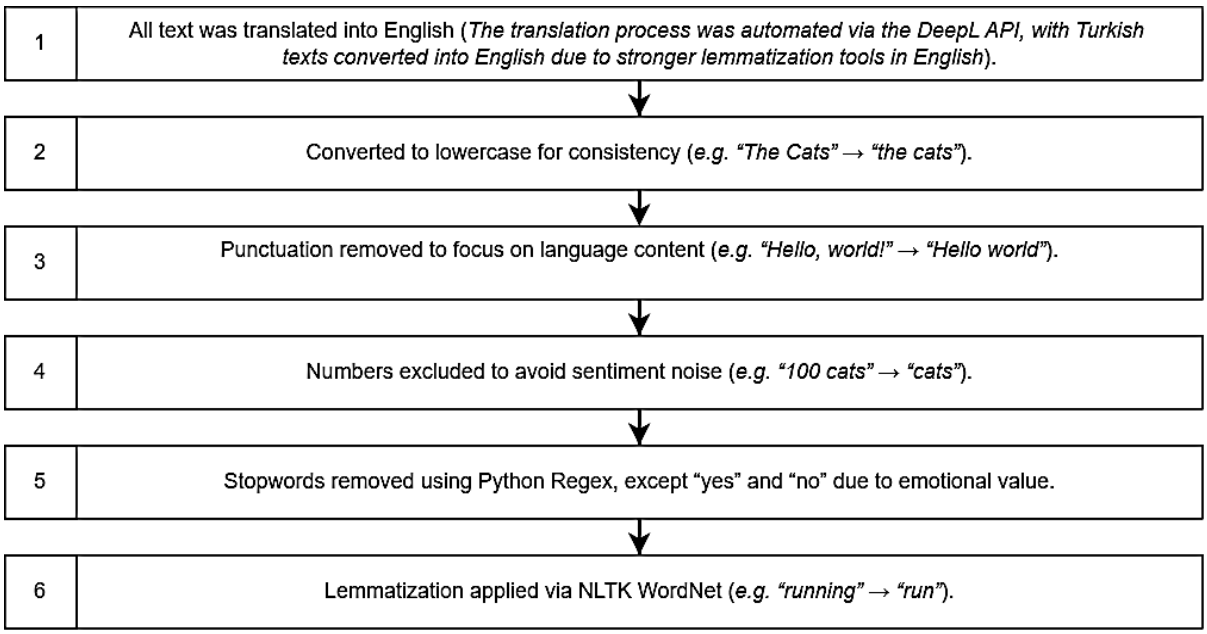
Research data were analyzed using NLP and machine learning. After text preprocessing, sentiment analysis employed TF-IDF, Word2Vec, and FastText with classifiers to estimate student emotions. Topic modeling explored opinions on distance learning. The process is explained in detail under the following subheadings.

Data Preprocessing

Data preprocessing involves a series of operations performed on a dataset prior to the sentiment extraction phase. Our preprocessing steps, illustrated in Figure 1, aimed to prepare the data for model training and testing, ensuring the model would be both understandable and reliable.

Figure 1

Data Preprocessing Steps



Sentiment Analysis

Sentiment analysis is a natural language processing technique used to determine the emotional content within texts (Devika et al., 2016). It aims to ascertain whether a piece of text contains positive, negative, or neutral sentiments. In this work, we tackled sentiment classification with a dataset of 4,439 user-generated opinions, of which only 1,961 instances were manually labeled by experts (positive: 858 vs. negative: 1,103). Given the scarcity of labeled data, we adopted a semi-supervised learning approach to leverage the remaining unlabeled examples.

After preprocessing, phrase vectorization was applied to convert text into numerical formats suitable for machine learning. Common methods include TF-IDF, Word2Vec, and FastText. TF-IDF generates vectors based on word frequency across documents. These methods do not consider word order or meaning—they only focus on word frequencies (Khanna & Coumans, 2023). Word2Vec, based on neural networks, also reflects the context and meaning of words into vectors (Garcia, 2021). FastText, on the other hand, converts words into vectors and also uses the sub-parts (n-grams) within the word (Patil et al., 2023).

After vectorization, the model was trained using five classification algorithms: naive Bayes, logistic regression, support vector machines (SVM), decision tree, and random forest (RF). Combined with three vectorization methods, this means 13 models were tested. These algorithms aim to accurately and efficiently categorize data into predefined classes (Alweshah, 2019). Classification algorithms are used in a wide range of applications, from text classification to medical diagnosis, and from image recognition to financial matters. Traditional machine learning models, such as k-nearest neighbors (k-NN), naive Bayes, SVM, decision trees, and RF, have historically been successful in sentiment analysis (Rodríguez-Ibánez et al., 2023). These models rely on statistical algorithms and predefined features to classify text sentiments such as positive, negative, or neutral.

In the following section, the classification algorithms used in this study are discussed in detail.

Topic Modeling

After the sentiment analysis, unlabeled sentiments were predicted with the best model, and the opinions in two categories (advantage and disadvantage of distance education) were analyzed separately with topic modeling. Topic modeling is a NLP method that aims to analyze large amounts of text data to uncover hidden thematic patterns (Gerlach et al., 2018). While topic modeling provides a systematic and scalable way to identify themes in large datasets, it differs from traditional qualitative thematic analysis, which relies on manual coding and interpretive validation. In this study, themes were derived directly from model outputs and researcher interpretation, without triangulation from additional qualitative methods. This is a particularly useful way of analyzing large amounts of student opinion. Topic modeling has been successfully applied in many areas, and topics in large datasets have been revealed (Ozyurt & Ayaz, 2022).

While determining the ideal number of topics and the modeling algorithms for topic modeling analysis, various metrics were used in addition to the interpretability-explainability of the topics that emerged. In this study, while defining the ideal model, LDA and CorEx topic modeling algorithms were tested by calculating coherence (C_V), homogeneity, precision (micro), recall (macro), and F1 score values. The algorithms were run separately to create topics between one and twenty and graphed (Figures 2 and 3), and both the algorithm and the number of topics were decided in this way. While consistency measures the degree of semantic similarity between high-scoring words on a topic, it is a frequently used metric

for the balance between human interpretability and computational efficiency (Mifrah & Benlahmar, 2020). Homogeneity evaluates the similarity of documents assigned to the same topic and the harmony of topics in different documents and is very important to ensure that topics are not overly broad or ambiguous (Amaro & Bacao, 2024). Micro precision assesses model accuracy per topic, macro recall measures capturing all relevant terms, and the F1 score balances both measures (Virtanen & Girolami, 2019).

There are many algorithms developed for this method, notably LDA. Although LDA is commonly preferred (Vayansky & Kumar, 2020), it is noted that it does not always provide the best result. It is particularly insufficient in short texts (Özyurt & Akcayol, 2020) and in revealing the relationships between words. Correlation explanation (CorEx) is also used for topic modeling. CorEx offers a compelling alternative to LDA by leveraging an information-theoretic framework that learns maximally informative topics without the need for detailed assumptions or complex hyperparameter specification (Gallagher et al., 2017). This approach not only simplifies model complexity but also facilitates the incorporation of human input through anchor words, enhancing topic separability and representation with minimal intervention.

Results

In this section, results are organized according to the research questions. The findings related to the first research question, being foundational, were prioritized and used to inform the analysis of the findings for the second research question.

RQ1: How are middle school students' perceptions of remote education distributed across positive and negative sentiment?

The model performances of the vectorization and classification algorithm pairs are presented in Table 1. When the performance data in Table 1 is examined, the SVM model trained with TF-IDF feature extraction stands out as the most successful method in terms of both accuracy (85.2%) and F1 score (85.1%).

Table 1

Model Performances of Vectorization and Classification Algorithm Pairs

Vectorization	Model	Accuracy	F1 Score
TF-IDF	Naive Bayes	0.811545	0.807016
TF-IDF	Logistic regression	0.823430	0.820996
TF-IDF	**SVM**	**0.852292**	**0.851196**
TF-IDF	Decision tree	0.784380	0.785191
TF-IDF	Random forest	0.825127	0.825088
Word2Vec	Logistic regression	0.723260	0.718766
Word2Vec	SVM	0.735144	0.727256
Word2Vec	Decision tree	0.650255	0.650678
Word2Vec	Random forest	0.760611	0.755425

Vectorization	Model	Accuracy	F1 Score
FastText	Logistic regression	0.672326	0.648586
FastText	SVM	0.563667	0.406379
FastText	Decision tree	0.602716	0.601008
FastText	Random forest	0.706282	0.696931

Note. The most successful pairing—TF-IDF and SVM—is shown in bold. SVM = support vector machines.

When the labeled data was estimated with the machine learning model created with this pair, it was observed that a significant portion of students stated there was no advantage to the advantage question in the preprocessing step, and similarly, they gave answers to the disadvantage question stating there was an advantage. The results are consistent with our observation; since students expressed their opinions without focusing on the question heading, some positive responses under "disadvantage" aligned with negative responses to "advantage," and vice versa.

A summary presented in Table 2 gives the total number of opinions, positive comments, and negative comments for each question. Although a total of 2,217 opinions were stated for the advantage, 509 of those were negative, i.e., saying there was no advantage. Similarly, 170 of the 2,222 total disadvantage opinions did not contain positive, i.e., disadvantage-oriented opinions.

In sum, positive opinion was 1,878 and negative opinion was 2,561.

Table 2

Summary of Total, Positive, and Negative Opinions for Each Question

Question	Total comments *N*	Positive comments *N*	Negative comments *N*
What is the advantage of distance learning?	2,217	1,708	509
What is the disadvantage of distance learning?	2,222	170	2,052

The results indicate that many students questioned or did not perceive any advantages of distance education, reflecting a skeptical view of its benefits and a clearer, more negative attitude toward its drawbacks.

RQ2: What latent themes underlie middle school students' perceptions of distance education?

CorEx and LDA were compared to identify latent topics, evaluated by coherence, homogeneity, precision, recall, and F1 (see figures 2 and 3). Detailed findings are explained in the following sections.

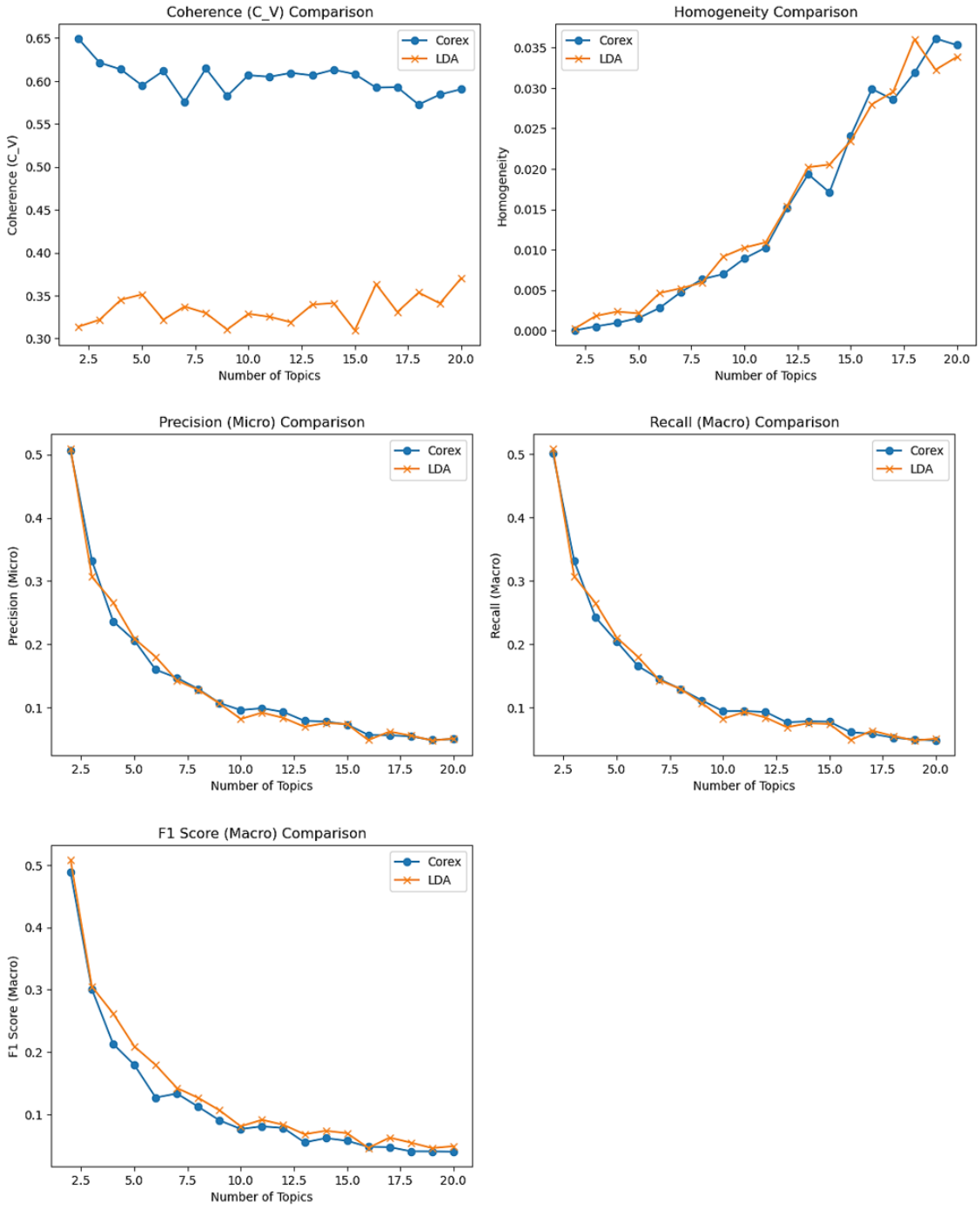
RQ2a: Which themes emerge from students' positive opinions?

Figure 2 shows CorEx achieved optimal coherence (~0.60) with six to eight topics, exceeding the meaningful threshold of 0.40, unlike LDA, which peaked early and dropped sharply (~0.33). While LDA showed slightly better F1 at two topics, its interpretability suffered. CorEx maintained better F1 at six topics (~0.13), balancing accuracy and readability. At high topic counts, both models show fragmentation, lowering insight value. CorEx-6 offered thematically distinct, interpretable topics,

aligning with qualitative goals. Thus, CorEx with six topics was chosen as optimal, and results are detailed in Table 3.

Figure 2

CorEx and LDA Performance Metrics for Positive Comments



Note. LDA = latent dirichlet allocation.

Table 3

Topics and Keywords for Positive Comments

Topic	Theme	Keywords
1	Continuity of education and health protection	education, distance, away, stay, continue, pandemic, period, technological, process, receive, class, close, virus, thank, tool
2	Time savings and physical comfort	advantage, think, school, don't, wear, tire, good, opportunity, sick, problem, road, situation, uniform, great, return
3	Noise-free, focused learning space	teacher, classroom, noise, sound, write, switch, environment, say, lesson, distract, voice, hear, turn, teach, image
4	Flexible scheduling and active participation	question, morning, ask, lecture, early, listen, want, solve, answer, test, read, book, comfortably, later, clothes
5	Self-regulated learning and information access	study, learn, time, information, home, like, health, little, access, spend, knowledge, covid, protect, work
6	Digital literacy and ease of learning	use, technology, make, eat, homework, way, efficiently, improve, assign, sense, hand, there's, share, sleep, easier

Using a 6-theme CorEx model (C_v = 0.446), the advantages of immediate distance learning for lower secondary students were clustered around six distinct themes listed in Table 3. Each theme is described below by triangulating the model's most important keywords with representative student quotes.

Theme 1: Continuity of Education and Health Protection

Statements such as "distance education [is] the best option in the conditions we live in" and "continue education without feeling risk due [to] the pandemic" illustrate that students framed remote teaching primarily as a means of maintaining academic progress while safeguarding health. The assured progression of the curriculum reduced uncertainty and fostered a sense of security.

Theme 2: Time Savings and Physical Comfort

Phrases such as "no need to wear [a] uniform" and "save time … no waste [of] time on the road" highlight gains in time management and bodily comfort. The removal of travel and dress codes redirected both physical and cognitive resources toward learning activities.

Theme 3: Noise-Free, Focused Learning Space

Excerpts such as "no noise environment—think calmly" show that the virtual classroom acted as a noise filter, improving audibility and concentration. Microphones and headsets facilitated clearer teacher–student communication and reduced peer distractions.

Theme 4: Flexible Scheduling and Active Participation

Students praised "not getting up early in the morning" and being able to "ask questions and get answers whenever [they] want," signaling that temporal flexibility empowered them to self-pace and actively engage with course content, aligning with principles of learner autonomy.

Theme 5: Self-Regulated Learning and Information Access

Comments such as "able to study at home and research" indicate strengthened self-regulation and unfettered access to digital information. The quieter home environment enabled deeper cognitive processing and individualized study routines.

Theme 6: Digital Literacy and Ease of Learning

Expressions including "learn to use technology tools" reveal that remote teaching served as a scaffold for digital literacy, while also enabling homework submission and making everyday needs (e.g., eating, breaks) more convenient, thus lowering affective and physiological barriers to learning.

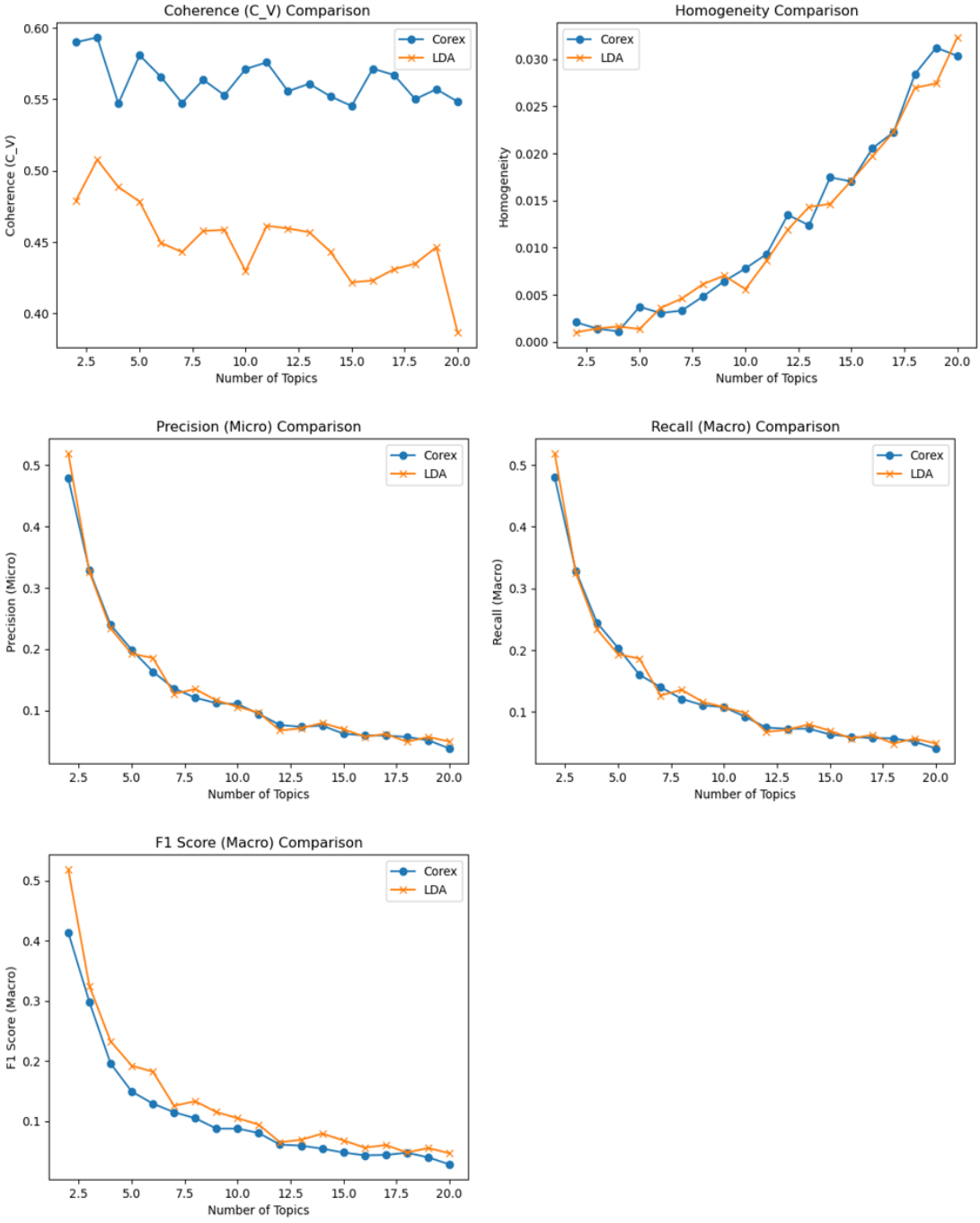
In summary, K–12 students in COVID-19 remote education saw it as safe, ensuring continuity, time savings, quietness, schedule flexibility, self-regulated learning, digital literacy, personalized education, and autonomy.

RQ2b: Which themes emerge from students' negative opinions?

For negative perceptions, Figure 3 confirms CorEx's superiority in interpretability. Its CV coherence remained high (~0.57–0.59) and stable between six to eight topics, outperforming LDA, which dropped sharply after $k = 2$ and stabilized near 0.44, with a practically significant 0.13-point gap. Although both models peaked in F1 at $k = 2$, the resulting topics were overly broad and lacked analytic depth. At six topics, however, CorEx maintained a double-digit macro-F1 (~0.12) with high coherence, whereas LDA's performance declined. Homogeneity rose beyond 10 topics, but recall collapsed (< 0.08), indicating excessive topic fragmentation. Thus, CorEx-6 was chosen to balance interpretability with classification performance for analyzing students' negative views.

Figure 3

CorEx and LDA Performance Metrics for Negative Comments



Note. LDA = latent dirichlet allocation.

Using the 6-theme CorEx solution (C_v = 0.396), students' negative perceptions of emergency distance learning coalesced into six interpretable disadvantage themes (Table 4). The theme labels were created by triangulating the most important keywords of the model with representative quotes and are explained in the following sections.

Table 4

Table of Topics and Keywords for Negative Comments

Topic	Theme	Keywords
1	Limited interaction and question resolution	question, ask, lesson, want, answer, enter, try, homework, live, attention, write, speak, difficult, teach, short
2	Screen fatigue and health strain	eye, look, screen, tire, health, hurt, deteriorate, computer, time, constantly, spend, long, affect, head, lot
3	Perceived low quality and inefficiency	education, distance, think, quality, learn, definitely, high, good, efficient, negative, opinion, fact, period, home, device
4	Audio/technical barriers to comprehension	teacher, voice, say, open, sound, work, hear, come, doesn't, explain, feel, throw, microphone, environment, classroom
5	Connectivity loss and cognitive breaks	understand, school, don't, subject, like, away, better, stay, fully, connection, issue, freeze, disconnection, face, productive
6	Access and participation constraints	attend, class, connect, low, people, participation, opportunity, attendance, make, unable, limit, end, wrong, broken, lecture

Theme 1: Limited Interaction and Question Resolution

Excerpts such as "sometimes may not be able to enter [the virtual classroom], may not get the answer we want" show that synchronous platforms often fail to replicate classroom dialogue, leaving questions unresolved and diminishing perceived teacher support.

Theme 2: Screen Fatigue and Health Strain

Students repeatedly mentioned "eye hurt … front of computer constantly" and "neck, back ache," indicating that the digital format imposed physiological costs that accumulated over long sessions.

Theme 3: Perceived Low Quality and Inefficiency

Phrases including "cannot get efficiency … distance education insufficient" reveal a widespread belief that remote delivery diluted instructional quality, especially for numeracy-heavy subjects.

Theme 4: Audio/Technical Barriers to Comprehension

Students complained that "sound goes, cannot hear teacher" and that overlapping microphones confused discourse, reflecting signal-to-noise problems well documented in other synchronous e-learning studies.

Theme 5: Connectivity Loss and Cognitive Breaks

Reports of "freeze, disconnection, throw us out" illustrate that unstable networks fractured cognitive continuity, forcing learners to reconstruct content gaps and eroding learning efficacy.

Theme 6: Access and Participation Constraints

Statements such as "no technological device … cannot attend" and "low participation" emphasize structural inequities: limited hardware, bandwidth, and moderated turn-taking all curtailed active engagement.

In summary, K–12 students in remote education during COVID-19 faced limited teacher interaction, delayed feedback, perceived low-quality instruction, especially in math, physical discomfort, technical disruptions, and inequities.

Discussion

This study examined middle school students' distance learning experiences and their emotional responses during the COVID-19 pandemic through sentiment analysis and topic modeling techniques. We found that students' comments generally contained negative sentiment. Additionally, students emphasized certain advantages of distance learning, such as ensuring continuity of learning while protecting individual health, saving time, and enabling flexible scheduling. However, student perspectives in general drew attention to disadvantages of distance learning, such as limited interaction and problem-solving opportunities, screen fatigue, and negative effects on health. Moreover, they emphasized the kind of structural barriers that disrupt the learning process, including technical problems, connection interruptions, and access issues. Reduced feedback and collaboration maps to weaker CoI teaching and social presences, while rigid formats and interruptions raise transactional distance; for early adolescents, limited self-regulated learning (SRL) scaffolds magnify disengagement (Garrison, 2016; Garrison et al., 2000; Moore, 1993; Pintrich, 2004; Zimmerman, 2002).

The sentiment analysis results show that students predominantly focused on negative characteristics may be related to the adverse situation created by extraordinary changes in their lives, such as the COVID-19 pandemic and the resultant widespread adoption of distance learning for students accustomed to face-to-face education. Indeed, studies have demonstrated that students experienced problems with engagement and motivation in distance learning courses during this period (An et al., 2023). Nevertheless, positive views were also found at a considerable level. The fact that distance learning served as the fundamental method for educational continuity during these challenging times may have been the primary basis for these views (Leech et al., 2022). Because patterns arose under ERT, implications would be crisis-contingent rather than general to fully designed online learning.

The topic modeling results reflecting positive views demonstrate that students experiencing distance learning for the first time at the K–12 level evaluated this process as a safe solution that ensured uninterrupted continuation of education under pandemic conditions, while simultaneously discovering unexpected advantages such as time savings, noise-free learning environments, and flexible scheduling. This finding aligns with previous studies that have emphasized the value of digital solutions in ensuring learning continuity during crisis periods (Datta & Nwankpa, 2021). Factors such as the elimination of commuting time, the absence of uniform requirements, and learning in the home environment enabled students to experience the educational process more comfortably, as was shown also in Tuguic & Bilan (2023). Having a quiet learning environment free from distracting elements and the opportunity to

follow courses with a flexible schedule suitable to their own learning pace were positive attributes of distance learning for the students. However, these findings differ from other literature. For example, there are studies that have emphasized that some students in distance learning encounter more distracting elements since they are usually in home environments, and teachers experience difficulties in ensuring student participation in lessons (Kadirhan & Sat, 2024). This difference may have emerged because the teachers in our study were able to reduce problems such as natural noise by using technological tools. Indeed, studies have reported that noise in face-to-face classroom environments affects student attention and motivation (Caviola et al., 2021). The opportunity for self-regulated learning and easy access to information, along with high levels of digital literacy and the simplicity of the learning process, were highlighted by some students who said they had the opportunity to develop their digital skills (Naidu, 2019). Still, such advantages may be unevenly distributed; skills in dealing with second- and third-level digital-divide factors and the translation of use into academic benefit can shape who realizes these gains (van Dijk, 2006).

On the other hand, negative views were quite diverse and could be classified at structural, pedagogical, and individual levels. Students specifically indicated that synchronous platforms could not reflect classroom dialogue due to limited interaction. Questions remained unanswered, and teacher support decreased. This situation demonstrates that the limited student-teacher interaction in distance learning can negatively affect learning (Dokuchyna, 2023). In CoI terms, fragile social and teaching presences explain lower belonging and participation; in theory of transactional distance (TDT) terms, low dialogue relative to structure widens transactional distance, risking misunderstanding and withdrawal (Garrison, 2016; Moore, 1993). Physical discomforts such as eye fatigue, headaches, and neck and back pain caused by prolonged screen use among students emerged as a significant problem. This situation indicates that online education needs to be redesigned from ergonomic and health perspectives (Upadhyay et al., 2021). Students' widespread beliefs that distance learning is particularly inefficient in mathematics and other courses with a lot of numeric content and reduces teaching quality emerged as a significant problem. These views support the necessity of strengthening student and teacher support to improve quality in distance learning, especially at the K–12 level (Martin et al., 2022). Technical problems, particularly audio quality issues, chaos created by overlapping microphones, and audio signal problems resulting in teachers being inaudible, make it difficult for students to understand lesson content; while connections drop and freeze, and platform disconnections fragment cognitive continuity, there is a need to reconstruct content gaps and strengthen learning effectiveness (Nowak & Watt, 2022). Finally, structural inequalities such as device shortages, insufficient bandwidth, and sequential speaking rights prevent students' active participation in lessons, indicating that the digital divide deepens learning inequalities (Solano-Gutiérrez, 2024). Addressing these barriers calls for actionable steps. For example, educators could: (a) balance structure with frequent, low-friction Q&A channels (TDT); (b) restore presence via predictable micro-feedback and brief synchronous check-ins (CoI); (c) scaffold SRL with weekly planners and progress prompts; and (d) couple access initiatives with digital-literacy supports to mitigate second and third-level digital divide effects (van Dijk, 2006).

This study demonstrates that NLP, sentiment analysis, and topic modeling can effectively analyze middle school students' unstructured opinions, enabling meaningful insights from large datasets. It extends prior work in distance learning (Borazon et al., 2024) and applies proven techniques from news and literature analysis (Gurcan et al., 2021; Lee et al., 2023) to the education domain. Given the short, grammar-sparse nature of K–12 texts, CorEx provided interpretable themes; a lightweight transformer

baseline (e.g., BERT) could contextualise performance without inflating word count and is reserved for future work.

Conclusion

In conclusion, the COVID-19 pandemic brought both notable benefits and serious challenges to middle school students' first experiences with online education. While students saw it as a safe way to maintain learning and gain autonomy, they also faced issues such as limited interaction, physical strain, and perceived quality loss in education. The use of NLP, machine learning, and topic modeling proved effective in quickly analyzing emotional feedback and insights.

In terms of implications for practice, we recommend short synchronous sessions with structured Q&A, regular micro-feedback, low-bandwidth/recorded alternatives, device–Internet support, and ergonomics-aware screen-time limits. Strengthening teacher training and promoting equity-focused initiatives will be essential to ensure that digital education becomes both inclusive and developmentally appropriate.

Limitations and Future Studies

Future research should extend the NLP-based approach to subgroup analyses (e.g., age, gender, region, and student background) to yield more fine-grained insights and guide context-sensitive policy and practice. The use of convenience sampling may introduce bias and limit generalizability; however, due to the extraordinary conditions of COVID-19, this approach was the most feasible, and future research should aim to employ more representative sampling strategies. While traditional qualitative methods can capture richer nuance, they are difficult to apply at this scale; NLP enables large-scale analysis and, when integrated with learning management systems, can provide systematic and timely insights despite some loss of subtlety. Future work should therefore combine NLP with in-depth qualitative approaches (e.g., manual coding, interviews) to balance breadth and depth and to triangulate model outputs. Finally, although CorEx was selected for short-text interpretability, future research could compare it with other state-of-the-art methods (e.g., BERT) and evaluate which approaches yield the most effective results in capturing nuance.

Acknowledgements

This article has benefited from the use of ChatGPT 5 (Extended) to polish the language after it was composed by the author. The authors meticulously reviewed and approved the final version of the content and assume full responsibility for the work.

References

Adams, D., Cheah, K. S. L., Thien, L. M., & Md Yusoff, N. N. (2024). Leading schools through the COVID-19 crisis in a South-East Asian country. *Management in Education*, *38*(2), 72–78. https://doi.org/10.1177/08920206211037738

Akhmedov, F., Abdusalomov, A., Makhmudov, F., & Cho, Y. I. (2021). LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS) model. *Applied Sciences, 11*(23), Article 11091. https://doi.org/10.3390/app112311091

Albert, A. (2024). From classrooms to confinement: Academic challenges faced by secondary school children in Kyamuhunga Sub-County during COVID-19. *International Journal of Innovative Science and Research Technology, 9*(9), 3206–3211. https://doi.org/10.38124/ijisrt/IJISRT24SEP1074

Al Husaeni, D. F., Haristiani, N., Wahyudin, W., & Rasim, R. (2022). Chatbot artificial intelligence as educational tools in science and engineering education: A literature review and bibliometric mapping analysis with its advantages and disadvantages. *ASEAN Journal of Science and Engineering*, *4*(1), 93–118. https://doi.org/10.17509/ajse.v4i1.67429

Alweshah, M. (2019). Construction biogeography-based optimization algorithm for solving classification problems. *Neural Computing and Applications*, *31*(10), 5679–5688. https://doi.org/10.1007/s00521-018-3402-8

Amaro, A., & Bacao, F. (2024). Topic modeling: A consistent framework for comparative studies. *Emerging Science Journal*, *8*(1), 125–139. https://doi.org/10.28991/ESJ-2024-08-01-09

An, X., Chai, C. S., Li, Y., Zhou, Y., & Yang, B. (2023). Modeling students' perceptions of artificial intelligence assisted language learning. *Computer Assisted Language Learning*, *38*(5-6), 987–1008. https://doi.org/10.1080/09588221.2023.2246519

Baral, S., Botelho, A. F., Santhanam, A., Gurung, A., Erickson, J., & Heffernan, N. T. (2023). Investigating patterns of tone and sentiment in teacher written feedback messages. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial intelligence in education. Posters and late breaking results, Workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky* (CCIS, Vol. 1831, pp. 341–346). Springer. https://doi.org/10.1007/978-3-031-36336-8_53

Bittencourt, I. I., Chalco, G., Santos, J., Fernandes, S., Silva, J., Batista, N., Hutz, C., & Isotani, S. (2024). Positive artificial intelligence in education (P-AIED): A roadmap. *International Journal of Artificial Intelligence in Education*, *34*(3), 732–792. https://doi.org/10.1007/s40593-023-00357-y

Borazon, E. Q., Marques, S., & Saycon, D. R. (2024, November 14). E-learning adoption: A comparative analysis of public sentiments during COVID-19. *Information Technology for Development*, 1–29. https://doi.org/10.1080/02681102.2024.2423286

Caviola, S., Visentin, C., Borella, E., Mammarella, I., & Prodi, N. (2021). Out of the noise: Effects of sound environment on maths performance in middle-school students. *Journal of Environmental Psychology*, *73*, Article 101552. https://doi.org/10.1016/j.jenvp.2021.101552

Copur-Gencturk, Y., Li, J., Cohen, A. S., & Orrill, C. H. (2024). The impact of an interactive, personalized computer-based teacher professional development program on student performance: A randomized controlled trial. *Computers & Education*, *210*, Article 104963. https://doi.org/10.1016/j.compedu.2023.104963

Datta, P., & Nwankpa, J. K. (2021). Digital transformation and the COVID-19 crisis continuity planning. *Journal of Information Technology Teaching Cases*, *11*(2), 81–89. https://doi.org/10.1177/2043886921994821

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: A comparative study on different approaches. *Procedia Computer Science*, *87*, 44–49. https://doi.org/10.1016/j.procs.2016.05.124

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, *49*(1), 5–22. https://doi.org/10.1177/0047239520934018

Diz-Otero, M., Portela-Pino, I., Domínguez-Lloria, S., & Pino-Juste, M. (2023). Digital competence in secondary education teachers during the COVID-19-derived pandemic: Comparative analysis. *Education + Training*, *65*(2), 181–192. https://doi.org/10.1108/ET-01-2022-0001

Dokuchyna, T. (2023). Students' attitude to distance learning as a component of their mobility. *Pedagogy*, *95*(4), 439–448. https://doi.org/10.53656/ped2023-4.02

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence, 21*(1), 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Fang, B., Li, X., Han, G., & He, J. (2023). Facial expression recognition in educational research from the perspective of machine learning: A systematic review. *IEEE Access*, *11*, 112060–112074. https://doi.org/10.1109/ACCESS.2023.3322454

Gallagher, R. J., Reing, K., Kale, D., & Ver Steeg, G. (2017). Anchored correlation explanation: Topic modeling with minimal domain knowledge. *Transactions of the Association for Computational Linguistics*, *5*, 529–542. https://doi.org/10.1162/tacl_a_00078

Garcia, M. (2021). Embeddings in natural language processing: Theory and advances in vector representations of meaning [Review of the book *Embeddings in natural language processing: Theory and advances in vector representations of meaning*, by M. T. Pilehvar & J. Camacho-Collados]. *Computational Linguistics*, *47*(3), 699–701. https://doi.org/10.1162/coli_r_00410

Garrison, D. R. (2016). *E-learning in the 21st century: A community of inquiry framework for research and practice* (3rd ed.). Routledge. https://doi.org/10.4324/9781315667263

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances, 4*(7), Article eaaq1360. https://doi.org/10.1126/sciadv.aaq1360

Geulayov, G., Mansfield, K., Jindra, C., Hawton, K., & Fazel, M. (2024). Loneliness and self-harm in adolescents during the first national COVID-19 lockdown: Results from a survey of 10,000 secondary school pupils in England. *Current Psychology*, *43*(15), 14063–14074. https://doi.org/10.1007/s12144-022-03651-5

Gurcan, F., Ozyurt, O., & Cagitay, N. E. (2021). Investigation of emerging trends in the e-learning field using latent Dirichlet allocation. *The International Review of Research in Open and Distributed Learning, 22*(2), 1–18. https://doi.org/10.19173/irrodl.v22i2.5358

Hurling, A., Ballard, H., & Whitaker, A. (2024). The effects of COVID-19 on public secondary schools operating in various economic environments. A case study of schools in Metros South and Central Schools, Western Cape. *American Journal of Education and Practice*, *8*(5), 18–29. https://doi.org/10.47672/ajep.2471

Kadirhan, Z., & Sat, M. (2024). K–12 teachers' perceived experiences with distance education during the COVID-19 pandemic: A meta-synthesis study. *Turkish Online Journal of Distance Education*, *25*(3), 57–75. https://doi.org/10.17718/tojde.1320633

Khanna, A., & Coumans, J.-V. (2023). Distributed vector representations of neurosurgical admission notes predict discharge disposition. *Neurosurgery*, *69*(Suppl_1), 47. https://doi.org/10.1227/neu.0000000000002375_322

Lee, K., Kim, T.-J., Cefa Sari, B., & Bozkurt, A. (2023). Shifting conversations on online distance education in South Korean society during the COVID-19 pandemic: A topic modeling analysis of news articles. *The International Review of Research in Open and Distributed Learning, 24*(3), 125–144. https://doi.org/10.19173/irrodl.v24i3.7220

Leech, N. L., Gullett, S., Cummings, M. H., & Haug, C. A. (2022). The challenges of remote K–12 education during the COVID-19 pandemic: Differences by grade level. *Online Learning, 26*(1), 245–267. https://doi.org/10.24059/olj.v26i1.2609

Lin, H., & Chen, Q. (2024). Artificial intelligence (AI)-integrated educational applications and college students' creativity and academic emotions: Students and teachers' perceptions and attitudes. *BMC Psychology*, *12*(1), Article 487. https://doi.org/10.1186/s40359-024-01979-0

Lin, L., Zhou, D., Wang, J., & Wang, Y. (2024). A systematic review of big data driven education evaluation. *Sage Open*, *14*(2), Article 21582440241242180. https://doi.org/10.1177/21582440241242180

Lo, C. K., Cheung, K. L., Chan, H. R., & Chau, C. L. E. (2023). Developing flipped learning resources to support secondary school mathematics teaching during the COVID-19 pandemic. *Interactive*

Learning Environments, *31*(8), 4787–4805. https://doi.org/10.1080/10494820.2021.1981397

Marshall, M. N. (1996). Sampling for qualitative research. *Family Practice*, *13*(6), 522–526. https://doi.org/10.1093/fampra/13.6.522

Martin, F., Sun, T., Westine, C. D., & Ritzhaupt, A. D. (2022). Examining research on the impact of distance and online learning: A second-order meta-analysis study. *Educational Research Review*, *36*, Article 100438. https://doi.org/10.1016/j.edurev.2022.100438

Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocarranza-Prado, I., & Kreibel, D. (2023). Impact of artificial intelligence on assessment methods in primary and secondary education: Systematic literature review. *Revista de Psicodidáctica (English Ed.)*, *28*(2), 93–103. https://doi.org/10.1016/j.psicoe.2023.06.002

Mifrah, S., & Benlahmar, E. H. (2020). Topic modeling coherence: A comparative study between LDA and NMF models using COVID'19 corpus. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4), 5756–5761. https://doi.org/10.30534/ijatcse/2020/231942020

Mishra, R. K., Urolagin, S., Jothi, J. A. A., Neogi, A. S., & Nawaz, N. (2021). Deep learning-based sentiment analysis and topic modeling on tourism during COVID-19 pandemic. *Frontiers in Computer Science*, *3*, Article 775368. https://doi.org/10.3389/fcomp.2021.775368

Moore, M. G. (1993). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22–38). Routledge.

Mujahid, M., Lee, E., Rustam, F., Washington, P. B., Ullah, S., Reshi, A. A., & Ashraf, I. (2021). Sentiment analysis and topic modeling on tweets about online education during COVID-19. *Applied Sciences*, *11*(18), Article 8438. https://doi.org/10.3390/app11188438

Naidu, S. (2019). The changing narratives of open, flexible and online learning. *Distance Education*, *40*(2), 149–152. https://doi.org/10.1080/01587919.2019.1612981

Nowak, K. L., & Watt, J. H. (2022). Distance education during the COVID-19 shutdown: A process model of online learner readiness, experiences and feelings of learning. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), Article 293. https://doi.org/10.1145/3555184

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

Ozyurt, O., & Ayaz, A. (2022). Twenty-five years of education and information technologies: Insights from a topic modeling based bibliometric analysis. *Education and Information Technologies*, *27*(8), 11025–11054. https://doi.org/10.1007/s10639-022-11071-y

Özyurt, B., & Akcayol, M. A. (2020). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Systems with. Applications*, *168*, Article 114231. https://doi.org/10.1016/j.eswa.2020.114231

Patil, R., Boit, S., Gudivada, V., & Nandigam, J. (2023). A survey of text representation and embedding techniques in NLP. *IEEE Access*, *11*, 36120–36146. https://doi.org/10.1109/ACCESS.2023.3266377

Pelaez, A., Jacobson, A., Trias, K., & Winston, E. (2022). The Turing teacher: Identifying core attributes for AI learning in K–12. *Frontiers in Artificial Intelligence*, *5*, Article 1031450. https://doi.org/10.3389/frai.2022.1031450

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Radwan, E., Radwan, A., & Radwan, W. (2020). The role of social media in spreading panic among primary and secondary school students during the COVID-19 pandemic: An online questionnaire study from the Gaza Strip, Palestine. *Heliyon*, *6*(12), Article e05807. https://doi.org/10.1016/j.heliyon.2020.e05807

Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, *223*, Article 119862. https://doi.org/10.1016/j.eswa.2023.119862

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sandvik, L. V., Smith, K., Strømme, A., Svendsen, B., Aasmundstad Sommervold, O., & Aarønes Angvik, S. (2024). Students' perceptions of assessment practices in upper secondary school during COVID-19. *Teachers and Teaching*, *30*(7–8), 918–931. https://doi.org/10.1080/13540602.2021.1982692

Scully, D., Lehane, P., & Scully, C. (2021). "It is no longer scary": Digital learning before and during the COVID-19 pandemic in Irish secondary schools. *Technology, Pedagogy and Education*, *30*(1), 159–181. https://doi.org/10.1080/1475939X.2020.1854844

Solano-Gutiérrez, G. A. (2024). La tecnología en la educación a distancia: Revisión de progresos y obstáculos a supercar [Technology in distance education: A review of progress and obstacles]. *Revista Científica Zambos*, *3*(2), 48–73. https://doi.org/10.69484/rcz/v3/n2/17

Tomaszewski, W., Zajac, T., Rudling, E., Te Riele, K., McDaid, L., & Western, M. (2023). Uneven impacts of COVID-19 on the attendance rates of secondary school students from different socioeconomic backgrounds in Australia: A quasi-experimental analysis of administrative data. *Australian Journal of Social Issues*, *58*(1), 111–130. https://doi.org/10.1002/ajs4.219

Tuguic, L. A., & Bilan, H. P. (2023). College education students' learning experiences on the advent of online distance education in the Philippines: A phenomenological study. *The Qualitative Report*, *28*(7), 1869–1879 https://doi.org/10.46743/2160-3715/2023.5644

Tzimiris, S., Nikiforos, S., & Kermanidis, K. L. (2023). Post-pandemic pedagogy: Emergency remote teaching impact on students with functional diversity. *Education and Information Technologies*, *28*(8), 10285–10328. https://doi.org/10.1007/s10639-023-11582-2

Uçar, S.-Ş., Aldabe, I., Aranberri, N., & Arruarte, A. (2024). Exploring automatic readability assessment for science documents within a multilingual educational context. *International Journal of Artificial Intelligence in Education*, *34*(4), 1417–1459. https://doi.org/10.1007/s40593-024-00393-2

Upadhyay, H., Juneja, S., Juneja, A., Dhiman, G., & Kautish, S. (2021). Evaluation of ergonomics-related disorders in online education using fuzzy AHP. *Computational Intelligence and Neuroscience*, *2021*(1), Article 2214971. https://doi.org/10.1155/2021/2214971

van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. *Poetics, 34*(4–5), 221–235. https://doi.org/10.1016/j.poetic.2006.05.004

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, Article 101582. https://doi.org/10.1016/j.is.2020.101582

Virtanen, S., & Girolami, M. (2019). Precision–recall balanced topic modelling. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 32, pp. 6747–6756). Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2019/hash/310cc7ca5a76a446f85c1a0d641ba96d-Abstract.html

Waheeb, S. A., Khan, N. A., & Shang, X. (2022). Topic modeling and sentiment analysis of online education in the COVID-19 era using social networks based datasets. *Electronics*, *11*(5), Article 715. https://doi.org/10.3390/electronics11050715

Xing, W., Li, H., Kim, T., Zhu, W., & Song, Y. (2025). Investigating the behaviors of core and periphery students in an asynchronous online discussion community using network analysis and topic modeling. *Education and Information Technologies*, *30*(5), 5561–5588. https://doi.org/10.1007/s10639-024-13038-7

Xu, H., Gan, W., Qi, Z., Wu, J., & Yu, P. S. (2024). *Large language models for education: A survey*. arXiv. arXiv:2405.13001. https://doi.org/10.48550/arXiv.2405.13001

Zarkadoulas, D., & Virvou, M. (2024). Emotional expression in mathematics e-learning using emojis: A gender-based analysis. *Intelligent Decision Technologies*, *18*(2), 1181–1201. https://doi.org/10.3233/IDT-240170

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Žitnik, S., & Smith, G. G. (2024). Automated analysis of postings in fourth grade online discussions to help teachers keep students on-track. *Interactive Learning Environments*, *32*(8), 4587–4612. https://doi.org/10.1080/10494820.2023.2204327