



Transitioning to Online Instructions and COVID-19 Response: A View from Mining Emergent College Students Discourse in Online Discussion Forum

Yiwen Lin¹ · Nia Nixon¹

Accepted: 14 May 2024 / Published online: 25 June 2024
© The Author(s) 2024

Abstract

The COVID-19 pandemic disrupted teaching and learning activities in higher education around the world. As universities shifted to remote instruction in response to the pandemic, it is important to learn how students engaged in learning during this challenging period. In this paper, we examined the changes in learners' social and cognitive presence in online discussion forums prior and after remote instruction. We also extracted emergent topics during the pandemic as an attempt to explore what students talked about and how they interacted with their peers. We extracted discussion forum posts between 2019 and 2020 from courses that have been offered repeatedly each term at a four-year university in the U.S. Our findings suggest that students exhibited higher social presence through increased social and affective language during remote instructions. We also identified emergent COVID-19 related discourse, which involved sharing personal experience with positive sentiments and expressing opinions on contemporary events. Our qualitative analysis further revealed that students showed rapport and empathy towards others, and engaged in active sense making of the pandemic through engaging in critical discourse. Our study sheds lights on leveraging discussion forum to facilitate learner experiences and building classroom community in online courses. We further discussed the potential for conducting large-scale computational linguistic modeling on learner discourse and the role of artificial intelligence in deriving insights on learning behavior at scale to support remote teaching and learning.

Keywords COVID-19 · Online Discussion Forum · Natural Language Processing

✉ Yiwen Lin
yiwenl21@uci.edu

¹ University of California, Irvine, USA

Introduction

The global COVID-19 pandemic significantly impacted all levels of education. In response to the quarantine order, universities and colleges shut down campuses and shifted to fully remote instructions to sustain teaching and learning activities (Strielkowski, 2020; Sun et al., 2020). Navigating learning in this unpredictable time can be challenging. Educators and learners had to deal with the stress and anxiety induced by the health crisis, while striving to engage in online courses and navigate digital learning systems (Pokhrel, & Chhetri, 2021). The lack of interaction and emotional support in physical isolation can expose students to greater challenges in online learning than regular times (Elmer et al., 2020). This underscores the need to understand what changes took place in the online learning environment that might potentially influence learner experience during the COVID-19 pandemic (Mishra et al., 2020). Studies have shown that the contingency shift to remote learning also the promoted digital revolution in higher education and creative solutions to cultivate an interactive and engaging environment in virtual learning (Strielkowski, 2020; Pokhrel & Chhetri, 2021).

Social and cognitive engagement are important factors for engaging in online learning (Ouyang & Chang, 2019). Online discussion forum plays a critical role in enabling communication and collaboration in remote settings and self-paced asynchronous courses. It promotes student-centric learning and peer interactions on specific topics with instructor guidance (Desai et al., 2021). Studies on Massive Open Online Courses (MOOCs) have found that discussion forums increase social presence and peer interactions, which enhances learners' psychological experience and learning outcomes. While discussion forums are an integral part of MOOC courses for question-answering and creating classroom communities due to the large student-to-teacher ratio, the utilization of discussion forums remained optional and is largely determined by instructors at accredited universities. In previous years, although LMS has been widely adopted by higher institutions for managing course materials, discussion forum remained to be a supplementary component to in-class lectures and discussions. One potential incentive for instructors to adopt discussion forums during remote instruction is that since discussion forums are already embedded in LMS, it requires lesser technological competency compared to other social facilitation tools. Studies suggest that discussion forums offer opportunities for information sharing and facilitating learners' soft skills (Suryaningsih, 2021), and support online assignments and questions (Mustadi et al., 2021). While there is reasons to believe the benefits discussion forum could bring regarding fostering social connectedness and learning community during the COVID-19 pandemic, there is a lack of understanding around the specific content of discourse and the changes of learners social and cognitive engagement in discussion forum prior to and after the contingency shift took place.

The goal of this study is to explore the emergent changes in student engagement in online discussion forum as a consequence of COVID-19 pandemic and fully remote instructions. Specifically, we aim to first provide a high level

view of how learners' participation, characterized by social and cognitive presence, in online discussion has changed, then compliment with evidence of what topics has emerged and whether students talk about pandemic in this formal learning environment. The findings aim to shed light on how teachers and students respond to a contemporary event that impacted teaching and learning, and how educational technology spaces such as discussion forum afforded these adaptations to happen. This study utilizes educational data mining techniques to harvest insights on teaching and learning at scale in a non-intrusive manner. We examine the topics and linguistic characteristics of students' posts in the discussion forum—within a subset of classes that consistently used discussion forum to control for course characteristics. The novelty of this study, compared to previous exploratory work on learner engagement in discussion forum, is our attempt to contextualize changes in learner interactions and examine emergent discourse attributed to the impact of the COVID-19 pandemic. The contribution of this work are three folds. First, we show how emergent machine learning methods can help us assess and monitor learning, and how educational technology spaces such as discussion forums afford different need of learning. As a methodological contribution, compared to survey and interview measures, we discuss the potential of a non-intrusive way to sample learning experience at scale. examine the impact of COVID-19 on learning behavior at scale. Second, we shed light on social, cognitive and affective aspects of learning at scale, offering both quantitative and qualitative to demonstrate not only what students talked about, but how they interacted with their peers. . Lastly, we offer discussions on implication of the study for on teaching and learning in higher education beyond the COVID-era.

The following sections are structured as follows. We first provide an overview of related work in educational data mining where researchers leverage large-scale educational discourse to gain insights into learning behavior. Then, we provide an account of research studying engagement in online discussion forums and identify constructs that are relevant for contextualizing teaching and learning in such environment. We describe the methods we used to quantify linguistic characteristics of learner discourse, and methods to extract topics relevant to COVID-19 within discussion forum during the early stages of remote learning. In the [results](#) section, we illustrate the change in sentiment during the academic year as well as the topics and sentiment characteristics associated with discourse around COVID-19. We present further interpretation of the results and implications in the [discussion](#) section.

Related Work

Educational Discourse and Text Mining

According to Vygotsky's (1978) sociocultural learning theory, learners bring their personal experiences into learning as they interact and build social relationships with others. From the psycholinguistic perspective, lexical expressions in writing and conversations disclose subtleties about people's internal thoughts, attitudes, and emotions (Pennebaker et al., 2003; Tausczik & Pennebaker, 2010). The value of peer

interaction in discussion forums is well documented (Gilbert & Dabbagh, 2005; Ziegler et al., 2014). Asynchronous online discussions promote articulation, reflection, social negotiation, and meaningful discourse that demonstrate critical thinking skills by relating course content to prior knowledge and experience in web-based or online learning environments. Language not only provides a window into learners' cognitive process but also characterizes the quality of classroom interactions (Bransford et al., 1999; Cazden, 1988). As the amount of student data generated online has increased exponentially during COVID-19, manually inspecting discussion content and their linguistic properties is no longer viable.

In the context of education, applied computational techniques are valuable tools to detect changes in response to the COVID-19 pandemic for the student population. A growing body of research has been centering on mining educational data to generate information that drive decision making in instructional activities (Romero, & Ventura, 2013). Emergent subcommunities such as learning analytics, educational data mining (EDM), artificial intelligence in education (AIED), and learning at scale (L@S) combine interdisciplinary knowledge to enhance our understanding of learner behavior and learner experience. In a time of crisis and sweeping changes taking place in higher education, being able to gather timely insights is particularly crucial to help teachers and students to come up with appropriate teaching and learning strategies. Computational techniques provide effective and efficient means to do so. For instance, student discourse provides critical information about learner experience and engagement and computerized text analysis allows us to delve into learner-generated text both in-depth and at scale.

To tackle the challenge of large unstructured educational text, NLP techniques, such as sentiment analysis and topic modeling, offer a means to process and obtain patterns from large-scale data (Romero & Ventura, 2017). The application of NLP in social media studies and product reviews in business is particularly prevalent, as an effective way to understand public discourse and sentiment at scale (Xue et al., 2020). In the field of education, a plethora of research has also demonstrated the benefits of text mining in addressing educational questions (Dowell & Kovanic, 2022; Lemay et al., 2021). For instance, studies have applied topic modeling to understand people's concerns and attitude towards online learning during COVID (Mujahid et al., 2021). Other researchers model learner-generated posts to surface insights around learning in formal and informal pedagogical environments. Chopra et al (2022) constructed topic chains by connecting semantically similar topics across months, demonstrating the temporal changes in learner discourse in the discussion forum during the pandemic. Sentiment analysis has been used widely to evaluate students' reflective writing (Chong et al., 2020) and student feedback (Altrabsheh et al., 2013; Lundqvist et al., 2020). It has also been used to track learner's emotional trajectory (Gkонтзис et al., 2017; Munezero et al., 2013; Neumann, & Linzmayer, 2021) and predict learning performance and satisfaction (Hew et al., 2020). Recently, Peng & Xu (2020) combined topic modeling and sentiment analysis to reveal significant differences in discourse behaviors from course review between completers and non-completers in MOOCs. They emphasized the importance of examining implicit discourse behavior (i.e., focused topics, topics' emotional tendencies and behavioral patterns) beyond explicit interaction features

(i.e. clickstream). To further understand sentiments expressed on specific topics, we can combine sentiment analysis with topic modeling to provide more contextualized details (Dolianiti et al., 2019a, b).

The contingent shift to online learning due to the COVID-19 pandemic brought huge opportunities for advancing research in understanding educational phenomena and searching for solutions to address challenges in higher education (Adedoyin & Soykan, 2023). Using learning analytics and EDM techniques (e.g., sentiment and topic modeling), we aim to provide insights into students' learning process and experience in Learning Management Systems (LMS).

Discussion Forum and Community of Inquiry Framework

Among literature on asynchronous discussion forums, the Community of Inquiry (CoI), which consists of three main components: cognitive, social, and teaching presence, is a framework to study forum activity (Garrison et al., 2001). According to the framework, social presence refers to the process when learners involve in interpersonal interactions and coordinate efforts with peers. This is further extended to reflect an individual's ability to identify within the community (Garrison, 2009). From the learning science perspective, social constructivists deemed the significance of interactions with peers and instructors that facilitate knowledge co-construction (Andrews, 2012). As such, facilitating social presence can be particularly crucial in distance education where facial expressions, body language, and auditory cues are lacking (Swan, 2010). In a way, social presence can be seen as 'the degree which a person is perceived as a "real person" and serve as a predictor of satisfaction in a computer-mediated environment (Gunawardena & Zittle, 1997). Research on social presence in online courses further established association to learner satisfaction and academic achievement (Joksimović et al., 2015; Kang et al., 2014), emphasizing the value of promoting social presence in online courses.

Cognitive presence represents higher-order thinking and constructing meaning through active reflection. Cognitive presence can be achieved when students link new concepts to past knowledge and reflect on the application of what they learned in class to real-life scenarios (Kilis & Yıldırım, 2018). Previously, research suggested that increased cognitive and social presence is beneficial to learning outcomes and psychological experience for online courses (Garrison & Arbaugh, 2007). A review also suggests that more purposeful interaction should be facilitated in future distance education (Abrami et al., 2011). However, there has been little empirical evidence to suggest how interaction patterns in asynchronous discussion forums changed during distance learning under the impact of COVID-19.

In response to COVID-19, courses that originally leveraged online discussion forums as an extension of classroom became more dependent on this digital environment for enabling student-to-teacher and peer-to-peer interactions when classes became fully online. Students' cognitive presence might also be seen in incorporating real-life events into critical discussions with classroom content. In addition, students might seek to connect with others more eagerly during remote instructions for the needs of social connection and belonging. Aside from using survey instruments

to directly measure CoI, studies have utilized computational linguistic program such as Linguistic Inquiry and Word Count (LIWC) to reflect socio-cognitive processes in discussion forum posts (Lin et al., 2020). In particular, cognitive process variable contain words that describe cognitive process and higher-order thinking (Moore et al., 2019; Pennebaker et al., 2015). Previous study has shown that LIWC-based analysis offers distinct proxies of cognitive presence (Joksimovic et al., 2014). The category of words social, which describes social content, and affect, which indicates emotional expression or self-disclosure, can be seen as signals of social presence (Ferreira et al., 2020).

Current Study

The current study seeks to examine the changes in discourse in online asynchronous discussion forums prior to and after the shift to remote learning and declaration of the global pandemic. We aim to shed light on the role of discussion forum from student posts. We take an exploratory approach to look at how social and cognitive presence in the forums have changed, and what emergent topics. We leveraged the alignment of COVID-19 development and instructional timeline to approximate the shock (i.e. the abrupt change in instructional activity) being introduced at the end of Winter quarter and prior to the start of Spring quarter. While the news around COVID-19 began to brew at the beginning of the year, the spread of the virus only gained more serious attention in the United States in March 2020 when the first case was discovered in the U.S. The outbreak was announced as a global pandemic by the end of March, which was also around the end of the Winter 2020 quarter (Cucinotta & Vanelli, 2020). As the government-issued stay-at-home order was in effect, the university took immediate action to shift to fully remote learning for the Spring 2020 quarter. Since the pandemic announcement coincides with the transition between Winter and Spring quarters, we consider that as a cutoff line to observe whether students' discourse in Spring is different from that which is observed in Winter and Fall. Figure 1 illustrates a timeline of the disease development with respect to instructional activities.

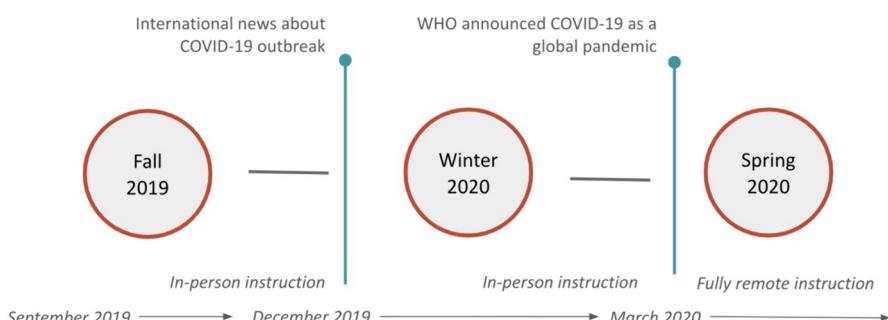


Fig. 1 Timeline of COVID-19 development and instructional activity

Specifically, we focus our investigation on the changes in social and cognitive presence in repeatedly offered courses. To control for course content and course level characteristics, we filtered out a set of courses that were offered every quarter and consistently used online discussion forum at the higher education institution where we obtained the data from. Additionally, we zoom in on the emergent discourse around COVID-19 and observed how social, cognitive, and affective components were manifested amongst those posts. Our current study investigates the following research questions:

- RQ1) How did social and cognitive presence in online discussion forum change before and after the transition to fully remote learning?
- RQ2) Whether topics around COVID-19 were present in the discussion forum?
- RQ3) What were the characteristics of COVID-related discussions in the forum?

We provide several hypotheses for the research questions. First, we hypothesize that social presence might increase under the assumption that discussion forum is used more heavily to facilitate social interactions to compensate for the lack of physical interactions. Research has shown that online discussion boards could effectively motivate student-student and student-instructor interactions (Bernard et al., 2009), especially in remote situations. In Ashokkumar and Pennebaker (2021), they found that language in a social forum indicating social connections and cognitive processes amplified following days when the growth rate of COVID-19 infections was higher, as a reflection of people's attempts to better understand and process the issues they were facing and seeking comfort from social ties when faced with threats. We were cognizant of the different nature between LMS forums and social media forums, where cognitive processes are inherently high in a learning context. We hypothesize that cognitive presence in student discourse might remain the same or slightly decrease, under the assumption that instructors might become more lenient and pose less complex problems for students to address in the discussions considering learners' limited mental capacity in this challenging time. Second, we expect some direct discussion on COVID-19 to be present in the discussion forum. Although forums tend to evolve around course content, COVID-19 and quarantine as collective experience should be detectable and might take up a non-trivial place in forum discussions. Lastly, we expect discourse around COVID-19 topics to be particularly high in social presence. Students might be more actively seeking social connections to cope with the challenging situation and build classroom community as a social buffer.

Method

Data

Our data was retrieved from an online learning management system at a large public university in the United States. Specifically, discussion forum posts naturally occurred during Fall 2019, Winter 2020, and Spring 2020 quarters across all courses

were obtained. In addition to forum posts, timestamp at the time a post was created, student ID, course and term information associated with the post, topic messages were also retrieved. We first removed posts that were not in English. We then further filtered messages based on the course level characteristics. Two inclusion criteria were used to determine whether a post would be selected. First, we aggregated forum posts within a course and filtered out courses that had fewer than 30 posts in a given quarter. Second, in order to eliminate course level differences, we retained courses that had been repeatedly offered and actively utilized discussion forum across all three quarters during the 2019–2020 academic year based on course code. The resulting courses are relatively consistent in curriculum design, course requirements, class size, and other course characteristics. The reason for restricting courses that have consistently used discussion forum is to reduce other potential confounding factor such as novelty effect when technological platform is first introduced to a class. As such, we can observe the changes in students' engagement in discussion forum in Spring more likely as a result of instructional or learning changes given the impact of COVID-19 pandemic. In total, 64 courses (22 in Fall, 20 in Winter, and 22 in Spring) remained in the dataset, comprised of writing seminars from the Humanities and English department. The courses are open to all students in the university and partially satisfy lower-division writing requirements. Since the writing requirement is a part of the General Education requirement that applies to all university students, students enrolled in these classes were predominantly first-year students. The course requirements and curriculum design are also relatively consistent across the quarters as foundational courses. We retrieved a total of 15,263 student posts from these classes. The total count of forum posts for the Fall, Winter, and Spring quarter were 4245, 4384, and 6634 respectively. The average posts per course were 192 in Fall, 219 in Winter, and 301 in Fall. Posts by instructors and teaching assistants were obtained but not included in the main analysis.

Participants

We obtained the administrative data on enrollment and matched with the student ID associated with the forum posts. In the dataset for analysis, a total of 1441 students enrolled in writing courses contributed to the forum posts. The enrollments by terms were 482 in Fall, 460 in Winter, and 532 in Spring. As students are required to take two lower-division writing courses, students may enroll in another course in a different quarter. As such, we observe a small the discrepancy in total number of enrollment and total number of unique students. According to the demographic data, we have 744 female and 697 male students in the data sample. Of those participants who reported their race and ethnicity (1384 out of 1441), 11.05% of the students were White, 3.32% were Black or African American, 58.60% were Asian, 26.51% were Hispanic or Latino, and less than 1% were identified as American Indian or Pacific Islander. Students enrollment status was primarily Freshman (96.6%) with a few transfer students.

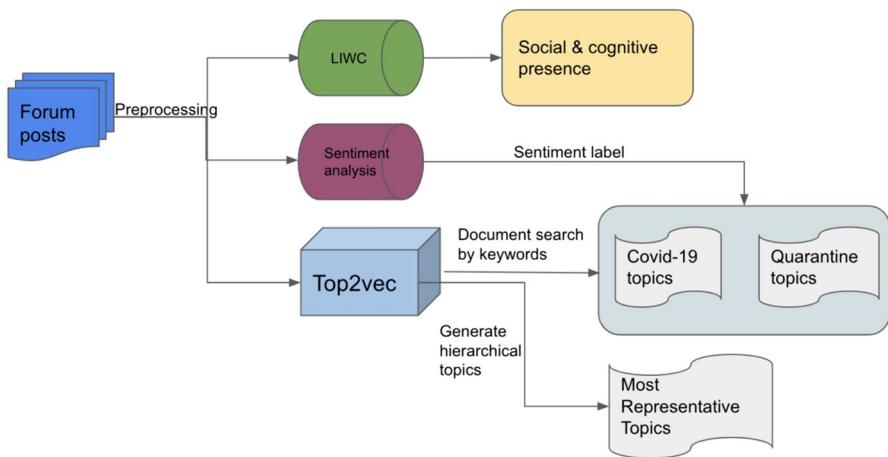


Fig. 2 Graphical representation of analysis workflow

Quantifying Online Presence and COVID-19 Presence in Discussion Forum

To characterize forum posts and identify COVID-19 relevant discourse in the entire forum, we employed several text mining methods. Figure 2 presents how we transform posts and generate linguistic measures across psychological and semantic dimensions for further quantitative analysis. To capture the social and cognitive presence, we used a lexicon-based computational linguistic program Linguistic Inquiry and Word Count (LIWC). We explain the rationale for the selection of LIWC properties for social and cognitive presence in the following section. We also performed a sentiment analysis to capture the positive or negative valence of discussion posts through sentiment scores. To explore the semantic dimension of learner discourse, we applied Top2Vec (Angelov, 2020), an unsupervised topic modeling technique to examine latent topics in the corpus. We performed a semantic search using keywords “covid” and “quarantine” to examine whether topics relevant to COVID-19 and quarantine emerged as a prevalent topic in the forum. We took a further look at the social and cognitive presence amongst quarantine and covid posts in order to examine whether COVID-19 related discussions were distinctively different from other forum discussions. We provide further details of each method in the following subsections.

Linguistic Inquiry and Word Count (LIWC)

Linguistic Inquiry and Word Count (LIWC) is an extensively validated dictionary-based tool for capturing psychological and linguistic properties in text (Tausczik, & Pennebaker, 2010). Each discussion forum post was considered a single document for analysis. We applied the LIWC2015 program to process each forum post, and the tool turned a score for each word category. LIWC captures a wide range of psychological dimensions and is capable of reflecting changes in people’s emotion,

Table 1 LIWC properties for social and cognitive presence

Construct	Linguistic variable	Subcategory	Examples
Social Presence	Social processes	Family, friend, female, male	They, buddy, neighbor, talk, girl
	Affective processes	Positive emotion, negative emotion, anxiety, anger, sadness	Happy, nice, sad, fearful, worried, love
Cognitive Presence	Analytic		
	Cognitive processes	Insight, Causation, Discrepancy, Tentative, Certainty, Differentiation	cause, know, think, how, should

cognition, and social connections during the early months of COVID development at the linguistic level (Ashokkumar & Pennebaker, 2021). In the context of online discussion forum, previous literature suggest that discussion posts that contain more affective, interactive and cohesive components demonstrate high social presence (Rourke et al., 1999; Hostetter, 2013). Ferreria et al. (2020) further suggest that several sociolinguistic indicators in LIWC can capture social presence automatically. As such, we selected LIWC categories that are representative of social presence, including social processes and affective processes.

Both social and affective processes have previously been considered features for constructing social presence (André et al., 2021). For cognitive presence, we included cognitive process and Analytic as indicators. Cognitive process captures higher-order thinking and critical thinking skills through words associated with causation, self-reflection, uncertainty, differentiation and so on (Moore et al., 2019). LIWC's cognitive processing score has been found to have high levels of predictive validity and has been used for automatic classification of cognitive presence empirically (Kovanović et al., 2016; Ferreira et al., 2020). Analytical thinking signifies formal and logical language which results from cognitive processes (Pennebaker et al., 2014). Table 1 shows the subcategory and example words of non-summary variables according to the LIWC2015 dictionary (Pennebaker et al., 2015). Note that analytical thinking is a summary variable that is calculated based on standardized scores from large comparison corpora and is a non-transparent variable in the dictionary.

Sentiment Analysis

Sentiment analysis is a common text mining technique to reveal people's opinions, attitudes and emotions toward an individual, events, or topic. In order to identify student opinions or attitude towards COVID-19 specific discussion, we created a sentiment score for each post using VADER (Hutto & Gilbert, 2014). VADER is a rule-based model for characterizing sentiment valences from written documents. We chose this method because VADER is interpretable, empirically validated, computationally efficient compared to BERT or other deep-learning based models, and a highly accurate tool to capture sentiment (Hilmy et al., 2019; Rääf et al., 2021). VADER has been widely applied across domains as well as in the educational

domain. While LIWC also has its categories representing positive and negative emotions, VADER is more sensitive and nuanced by incorporating lexical features such as emoticons, sentiment-related acronyms and initialisms, as well as commonly used slang with sentiment value (Hutto & Gilbert, 2014). Since discussion forum posts resemble the format of microblogs - the type of text VADER is trained on and most attuned to - this approach would be viable for our data. We used VADER within the Natural Language Toolkit (NLTK) library in Python to produce a normalized compound score for each post. The compound score calculates the sum of all lexicon ratings and returns a value from –1 (extremely negative) to 1 (extremely positive). The more a compound score is close to 1, the higher positivity is indicated in the text.

Top2Vec Modeling

To extract student discourse on COVID-19, we employed an unsupervised topic modeling technique named Top2Vec (Angelov, 2020). This modeling approach automatically detects topics present in text and determines the optimum number of topics. Top2Vec algorithm works on the assumption that many semantically similar documents are indicative of the underlying topic. It creates jointly embedded document and word vectors using Doc2Vec (Le & Mikolov, 2014), turns it into lower dimensional embedding of document vectors with dimension reduction technique (Uniform Manifold Approximation and Projection), and finds dense areas of documents. It then calculates the topic vector, which is the centroid of document vectors in the original dimension and calculates n-closest word vectors to the resulting topic vector. The original paper (Angelov, 2020) provides further algorithmic intuitions behind the model. Compared to traditional topic modeling such as LDA, an advantage of this unsupervised approach is that the model takes into account the semantic relationship of text and produces results at a more granular level. Instead of using bag-of-words (BoW) representation of documents which ignore the ordering and semantics of words, top2vec leverages joint document and word semantic embedding to find topic vectors. Moreover, the authors pointed out that the misconception that we commonly fall into topics are often thought of as discrete values (i.e. politics, science, art), when in reality topics can be further subdivided into many other subtopics. Unlike traditional topic modeling methods, such as LDA, top2vec does not rely on human input or parameter tuning during the training process.

While it is more efficient to use a universal sentence encoder and other pre-trained embeddings, we trained the Doc2Vec model from scratch due to the novelty of COVID-19 vocabulary. Between “fast-learn”, “learn”, and “deep-learn”, we set the parameter at “learn” to achieve a balance between speed and quality vectors. We consider each post as a document, and the entire collection of posts as corpus. Once the model is trained, we ranked the top 10 most prominent topics by topic size as well as topic words to interpret the topics. We further located specific topics using keywords. Top2Vec allows for topic and document searches by keywords. We did two separate searches by the keywords of “quarantine” and “covid / coronavirus” and extracted 5 topics that were most semantically relevant to each search term. Top2Vec returns a list of topics with an index and topic score to each topic. Topic

score illustrates the cosine similarity for each topic to the search keywords. We then retrieved a list of 50 words under each topic, ranked by word score (a cosine similarity score of the word to the topic). We generated word clouds for the top 5 most relevant topics to “quarantine” and “covid / coronavirus”. The size of the words in word cloud is determined by word score, signifying the importance of this word in the topic. For further analysis on posts, we retrieved 20 most relevant documents in each of the top 5 topics. For the posts retrieved from the five quarantine topics, we will refer to them as “quarantine posts” in the subsequent sections. For the posts retrieved from the five covid topics, we will refer to them as “covid posts” in the subsequent sections.

Statistical Analysis

To address our first research question, we compared the characteristics of forum posts in Spring and those that were posted in Fall and Winter. As depicted in Fig. 1, the alignment of COVID-19 development and instructional activities suggests that the beginning of the Spring quarter signals the transition to fully remote instruction. We generated a binary variable to differentiate whether a post was posted during remote instruction or prior to remote instruction. We conducted Welch two-sample t-test to compare the means of selected LIWC properties (as shown in Table 1) and sentiment scores between posts in Spring quarter and posts in Fall and Winter quarter.

To address the second research question, we turn to topic score, which indicates the importance of a given topic amongst other topics, to observe whether COVID-19 emerges as one of the most representative topics in the entire corpus. To address our third research question, we repeated the same analysis as in RQ1 to compare differences between covid-related posts and other posts.

Table 2 Enrollment, Course, and Posting Count by Quarter

	Enrollment (Number of students)	Total Posts	Total Number of Courses	Posting activity per course
Fall 2019	482	4245	22	192
Winter 2020	460	4384	20	219
Spring 2020 ^a	532	6634	22	301

The instructional period for each academic quarter is respectively Fall (September-December 2020), Winter (January-March 2020), Spring (March-June 2020)

^aA note for fully remote instruction

Table 3 Differences between Spring posts versus Fall/Winter posts on social and cognitive presence captured by LIWC variables. Welch test is reported because Levene's test indicated homogeneity of variances assumption was not met. Embolden signals $p < .05$

	Fall & Winter (in-person)		Spring (Fully remote)		<i>t</i>	<i>p</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Social Process	9.20	5.41	9.93	5.97	13514.52	-7.89	<0.001
Affect	5.08	3.30	5.57	4.56	11599.97	-7.45	<0.001
Analytic	72.15	23.98	70.26	25.17	13912.96	4.70	<0.001
Cognitive process	13.67	5.77	13.54	5.89	14125.57	1.35	0.02
Compound sentiment	0.28	0.74	0.33	0.71	14516.36	-4.97	<0.001
							-0.08

The instructional period for each academic quarter is respectively Fall (September-December 2020), Winter (January-March 2020), Spring (March-June 2020)

Results

Prior to reporting the results regarding our research questions, we would like to highlight some contextual information at the course and term level that would situate the interpretation of results. First, we observed an increase in overall posts numbers ($N_{Fall} = 4245$, $N_{Winter} = 4384$, $N_{Spring} = 6634$) in the discussion forums across all courses. The number of students who generated these posts were respectively 482, 460, and 532 in Fall, Winter, and Spring. The number of courses included in the dataset were 22, 20 and 22 in each term. We calculated the posting activities per class by using total posts in a given quarter divided by total number of courses in a given quarter. As shown in Table 2, we can see that the sum of forum posts per course increased from 192 in Fall and 219 in Winter, to 301 posts in Spring. This indicates a potential increase in active utilization of discussion forum for instructional purposes, or a more active engagement from students during remote instruction period.

Overall Online Presence

To our first research question, we found evidence on changes in social and cognitive presence in the discussion forum. Table 3 display the outcome of analysis based on LIWC. Our results suggest a significant increase in overall social presence in the online forums. Specifically, when we compared the linguistic characteristics of forum posts between the periods of in-person instruction (i.e. Fall and Winter) and fully remote instruction (i.e. Spring), two main linguistic categories (i.e. social, affect) that represent social presence showed significant differences. First, we found significantly more prominent social language in Spring ($M=9.94$, $SD=5.97$) than in Fall and Winter ($M=9.20$, $SD=5.41$). We also observe stronger affective language in the discussion forum during remote learning ($M=5.57$, $SD=4.56$) than the average level of affective language prior to remote learning ($M=5.07$, $SD=3.30$).

Secondly, consistent with our hypothesis, our results showed an overall similar level of cognitive presence before and after fully remote instructions, signaled by LWIC's cognitive processing. This suggests students engaged at similar cognitive reasoning level in their posts. However, we did find evidence for a decrease in analytical thinking marked by a decrease in analytical expression. Analytical thinking during Fall and Winter was ($M=72.15$, $SD=23.98$) and dropped to ($M=70.26$, $SD=25.17$) in Spring. While this does not mean students are less cognitively invested in the discussion assignments, it does indicate students were using lesser formal and less complex expression during remote instruction. This aligns with other literature suggesting the decrease in analytical expression and people's tendency to use more simple and straightforward expressions during COVID-19. We further elaborate on the implications of this result in the [discussion](#) section. With respect to the sentiment analysis results, spring quarter posts show significantly higher compound score, indicating an overall more positive sentiment in these posts. We conducted Cohen's d test estimate between-subjects effects for the grouped data. We interpreted Cohen's d effect sizes using a variation on Sawilowsky's extension

Table 4 List of top 10 topics and representative topic words

Topic	Topic Theme	Topic words
1	Course logistics	drafts, assignment, annotated, week, peer, feedback, papers, portfolio
2	Casual interactions	Quarantine, hello, freshman, classes, year, hobbies, major, fun, friends, excited
3	Writing	Quote, responses, paraphrasing, integration, concise, thorough
4	Social justice system	Prison, inmates, incarceration, parole, facilities, sentencing, reform
5	Music	Album, songs, music, artists, band, hiphop, fans, rock, beat, style
6	Race history and discrimination	Civil, criminal, racial, documentary, slavery, minorities, activist, justice, system
7	Feedback	Descriptive liked, interesting, responses, detailed, metaphor, relate, elaborate
8	Technology	artificial, intelligence, technology, machine, robots, advanced, algorithm, concern
9	Storytelling	Authenticity, emotion, characters, plot, story, vivid, writers, conflict, dilemma
10	Education issues	Schools, curriculum, funding, educational, university, student, achievement, programs

of Cohen's original scheme (Sawilowsky, 2009; Cohen, 1992; Windsor et al., 2019). According to this scheme, the effect sizes on social process ($|d| = 0.12$) and affective process ($|d| = 0.12$) were small, and the effect size on cognitive process ($|d| = 0.02$) and analytical process as well as sentiment analysis ($|d| = 0.08$) was very small.

COVID-19 Topic Analysis

To further illustrate emergent learner engagement in discussion forum during remote instruction, we examined the prominent topics across the time span of the Fall, Winter, and Spring quarter. First, we found that COVID-19 was among the main discussion. Table 4 shows a list of the top 10 topics and topic words within the topic. The topics are shown in the order of significance ranked by topic size. We can see that casual interaction related to quarantine is one of the more prominent topics detected in the corpus.

We found the 5 most relevant topics associated with “quarantine” is generally associated with casual interaction and students sharing personal interests or experience. We visualized the most semantically related topic words in these topics through word clouds in Fig. 3. A full list of the topic words (listed in order of importance to the topic) and topic scores (the cosine similarity for each topic to the search keyword) is available in the supplementary material. From the word clouds below, we can see that students were using pronouns for friends, family, and pets. There is also evidence to the types of activities and hobbies students do such as “beach”,

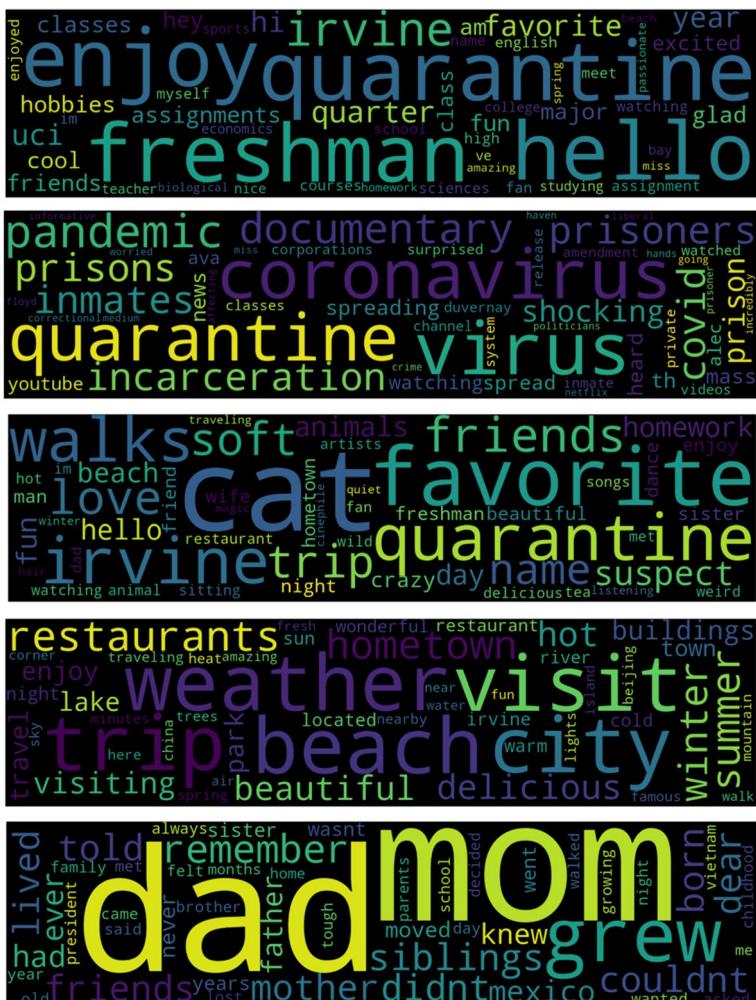


Fig. 3 Word cloud visualizations of top 5 topics extracted by keyword “quarantine”. Note. The size of words correlates with the importance of the words in given text. The larger and bolder the words are the more important they are to the topic

“walks”, and “trips”. We can also observe emotional words such as “shocking”, “surprised”, and “miss” that indicate the sharing of feelings.

Similarly, we extracted the top five most relevant topics retrieved with keyword “covid / coronavirus”. We found two of the topics overlapped with the previous search, which was unsurprising given the close semantic relatedness of the two search terms and indicates that they were used interchangeably in similar contexts. However, we did observe some differences in the other three topics compared to topics search results from “quarantine. The group of covid topics expands beyond casual interaction and has a stronger emphasis on discussions on societal events and

Table 5 Comparison between posts extracted by keyword “quarantine” and the rest of the discussion forum posts across LIWC variables representing social and cognitive presence.

		Non-quarantine		Quarantine		<i>df</i>	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Cognitive presence	Analytic	71.26	24.6	62.07	27.50	99	3.31	<.001	.35
	Cognitive process	13.64	5.90	12.28	5.55	99.43	2.43	.017	.24
Social presence	Social	9.54	5.69	9.71	7.59	98.70	-.22	>.05	-.02
	Affect	5.31	3.97	6.38	5.18	98.74	-2.06	=0.412	-.23
	Sentiment Compound	0.30	.73	0.47	.66	99.5	-2.57	=.012	-.24

Welch test is reported because Levene's test indicated homogeneity of variances assumption was not met. Embolden signals *p*<.05

opinion expression. For instance, in topic 66, the keywords indicate discussion on misinformation and fake news during COVID-19 outbreak; Topic 99 alludes to the social justice protests ignited by the death of George Floyd caused by police brutality in May 2020 and the black lives matter movement that happened as the coronavirus continued to spread. Topic 31 indicated timeline and discourse on politics and social media.

To obtain information on discourse characteristics around COVID-19 topics to address how students were talking about them, we extracted top 20 posts from the 5 most relevant topics associated with “covid / coronavirus” and “pandemic” separately, resulting in 100 posts “covid” post and 100 “pandemic” post. We then analyzed these posts both quantitatively and qualitatively. First, we compared the same LIWC properties as above (i.e. Analytic, social, cognitive process, affect) and compound sentiment of covid posts against non-covid posts using t-test. Our results suggest no significant difference that distinct covid posts from the rest. However, we found that covid posts are on average Next, we repeated the analysis on quarantine posts and non-quarantine posts. Our results suggest that quarantine posts show no significant differences in social process, but significantly higher affective process ($t(98.73) = -2.06$, $p = .042$, $d = 0.24$), lower cognitive process ($t(99.42) = 2.43$, $p = .017$, $d = 0.24$), and significantly lower analytical language ($t(99) = 3.32$, $p < .001$, $d = 0.35$) compared to other posts. Additionally, quarantine related posts show significantly higher compound score from sentiment analysis ($t(99.5) = -2.57$, $p = .012$, $d = -0.24$), indicating an overall more positive sentiment in these posts. Table 5 demonstrates the differences in social and cognitive presence between quarantine-related posts and non-quarantine-related posts. The effect sizes for affective and cognitive processes, compound scores ($|d| = 0.24$), and analytical process ($|d| = 0.35$) were considered medium according to the new scheme for effect size interpretation (Sawilowsky, 2009). There was no evident difference between the characteristics of posts extracted under covid topics.

To gain contextualized insights, we sampled several posts to provide more qualitative details amongst quarantine and covid related posts. We observed that students shared lived experiences, built connections, and showed empathy with peers. This discourse was associated with high positive sentiment valence. Under covid related topics, students frequently expressed gratitude, shared wishes and hope for others, and provided encouragement and support. For instance, one student acknowledged the challenges their peers were going through by sharing their own experience:

“I hope that quarantine is not as stressful as it initially was for you. My allergies began to act up around the same time as this whole mess so I totally understand you for that. I’m glad you made it back home in time and that you were able to reunite with your family.” In another example, one student expressed empathy to the respondent and established common grounds:

“...Thank you for sharing your online class experience and your time during quarantine. I’m also really worried about the online class as well, but I believe it’s going to be a good experience. I love to workout, so maybe we can have a virtual workout together during the quarantine time? haha anyways, nice to meet you, and hope you stay healthy! :)” This suggests high social presence amongst the posts and content-oriented around building social rapport.

In addition, we observed posts where students provide constructive support or feedback to peers. For example, a student wrote

“I really like your topic! I found it interesting and think it’s a great idea to relate the current situation to the first amendment. I think adding research to the pros and cons of social distancing would flow nicely with your work. Adding more context to COVID-19 would be a great addition too. You could explain what its is, the effects it has, and its severity...”

In another example, student emphasized the strengths in their peer’s work and pointed to actionable areas that they might improve on:

“I believe that your connection to the pandemic occurring right now and using COVID-19 as an example is really strong, and showcases how the problem is present right as we speak of it. Using this crisis and stating how it is changing how our communities are functioning is good, and I like how you touched upon the misrepresentation it may display. I believe that delving in a bit keeping into the COVID-19 situation in connection with the racial discrimination occurring against Asians right now would have been a really good example.”

Other posts centered on critical discussion along with other social issues. For example, one student discussed COVID-19 situation in relation to healthcare within the prison system:

“With COVID going on right now I focused on how we could help inmates and one of the main solutions right now is focusing on the safe release of inmates. However, I’ve found that that isn’t enough, because prisons would still be overcrowded and it wouldn’t prevent the spread of COVID that much.”

Another student discussed the racial discrimination against Asian American amplified by the pandemic:

“...As a result of COVID-19, there has been an increase in discriminatory actions, and the cause of this is due to authorities of higher power, such as politicians, giving rise to the association of COVID-19 with China, causing people to also associate China with all Asians and Asian-Americans in the US...”

These posts are characterized by positive sentiment, and suggest a type of discourse that demonstrates both high social presence and high cognitive presence.

Discussion

We found that there is an overall increase in social presence, characterized by increased social processes, affective processes, and positive valence in student discourse. Overall, this result demonstrated impact of COVID-19 pandemic on people's expression of emotions that were found in other online platforms (Monzani et al., 2021). However, unlike social media where negative sentiments were pervasive, the sentiments expressed in discussion forums were positive. This indicated that learners might have increased social connectivity in Spring with positive language, which was corroborated by our qualitative analysis. Upon further investigating the content of these posts, we found students established positive interactions by showing rapport and empathy amongst themselves. Interestingly, our results show little change in cognitive processes in learner discourse but a significant decrease in the analytical property. This may imply that while learners remained actively engaged in critical thinking and reasoning, it was expressed in a less analytical manner (i.e. language that are less formal, logical, and hierarchical). A possible explanation is that disruptive events could reduce analytic thinking and invoke more personal and informal language (Seraj et al., 2021). Monzani et al. (2020) corroborated with evidence that emotional tone and analytic thinking were lower in the first two months of the pandemic, which was characterized by uncertainty. Markowitz (2023) further suggests that one's interest and motivation to think was the driver behind reliable and robust connection between analytic thinking and cognition, not cognitive ability. Taken together, we may consider the change in analytical thinking signaling a drop of incentives to engage in linguistically complex manner, which prompts future examination on its association with students' psychological experience.

To our second research question, since the set of courses we focus on have discussion forum participation as a pedagogical component, the hypothesis is that the student discourse should remain consistent if nothing else is changed except courses moving to online format. However, we found the emergent COVID-19 discourse took up a non-trivial space in the corpus as indicated by topic rank. The presence of COVID-19 contents in discussion signals a few possible changes happened in teaching and learning. First, instructors facilitated interactions related to COVID-19 experiences. Second, the facilitation remained unchanged (e.g. “Introduce yourself

to class”), but learners organically leveraged discussion forum to foster peer connection or reasoning around COVID-19. For both possible explanations, we can infer that there is an emergent need for social connectivity during online instructions in Spring, perceived by instructors or inherently expressed by students. Interpreting with the results for RQ1, this increased social presence suggests there is an increase in socialization activity in learning during COVID-19.

To our third research question, we found that there are more nuances in the way students talked about COVID-19. While there were some overlaps between topics retrieved with “covid” versus “quarantine” (topic 103 and topic 1), the remaining topics were different. We can observe from the wordcloud visualization and also the topic words provided in the appendix, that the forum post content varies based on the search term. Quarantine posts centers on personal experience sharing and building social connections, whereas covid posts involve more objective discussion and sense making of this contemporary events. Quarantine posts showed significantly stronger affective language and more positive sentiment, compared to non-quarantine posts. Our qualitative analysis corroborates with this finding. After examining the details of quarantine posts, we found prominent themes of building resilience, sharing struggles and personal frustrations and showing social rapport. However, we did not observe a significant difference in social process between quarantine and non-quarantine posts. This suggests that quarantine messages were not a main contributor to high social process in spring. Instead, we could infer that the overall increase of social connectivity in spring forum came from students seeking to build social connections in various manners, not exclusive to messages related to quarantine. Interestingly, the results show that there is a significant drop in both cognitive and analytic processes for quarantine-related posts. This could indicate that the conversations on quarantine were less formal and cognition-oriented.

Measuring learners’ online engagement in discussion forum can provide valuable information to educators on adapting their pedagogical practice to learning needs or take necessary intervention. Overall, our findings underscore several the important role discussion forums can serve in remote instruction, specifically by affording classroom community building and social connectedness, and by engaging students in cognitive and analytical thinking. We identified posts that expressed social connection needs, which could imply the necessity of facilitating supportive classroom community. The increased social presence in a discussion forum during the Spring quarter expressed learners’ psychological need for social connections, and discussion forum served as a channel for them to interact with peers. We found student discourse that demonstrate rapport-building and emotional buffering that could be beneficial for coping with social isolation during quarantine. Discussion forum could also help instructors identify specific challenges students experience. For instance, students reported some of the biggest challenges during the lockdown include loneliness from social isolation and a lack of support in learning. Our study also has implications on application of educational technology, particularly on assessment. For instance, an overall decline in analytical language in Spring might signal students having limited capacity in processing complex subjects on top of the need for making sense of COVID-19 impacts. While many learning management systems are now equipped with

automatic assessment based on learners' language, we should be careful in interpretation and take into account of external factors.

There are several limitations to our study. First, our results regarding the differences in social and cognitive presence represented through LIWC between spring and non-spring quarters yielded small effect sizes. Considering psychological language markers tend to have modest effect sizes (Holtzman et al., 2019), we argue that the trend and pattern still provide meaningful information on how overall discourse has changed in the discussion forum prior to and after shift to remote instruction. The effect sizes for the comparison between quarantine-relevant posts and other posts were moderate, distinguishing quarantine posts as more positive and less analytic from non-quarantine content posts. Second, the sample of our courses was limited to a set of first-year writing seminars. This selection was intentional in order to ensure course level characteristics stay relatively salient, although it does not capture a wide range of disciplines and domains. We may also safely assume that teachers for these courses were familiar with facilitating discussion using online forums. As such, our result may not represent the activities in courses that were plunged online and used discussion forum for the first time. For courses that have been online and remained online courses during the pandemic, their discussion forum activities could appear differently as well. Despite the course level differences, our study mainly focuses on capturing the changes to courses that transitioned from in-person to fully-remote instruction due to the disruption of COVID-19, where discussion forum's role may have also shifted from an additional environment for out-of-classroom interactions, to a more centered space for online presence. Lastly, our current study focuses only on learner discourse and primarily peer interaction in the discussion forum. This might not paint a full picture of online learning because we did not investigate teacher's role or teacher-student interactions. However, by inferring from learner discourse, our study can indirectly reflect teacher facilitation. For instance, we may infer from learner discussion that instructor facilitated students to introduce themselves and share their experience in quarantine. Future studies should focus on teacher-student interactions to gain further insights into instructors' role in moderating and facilitating discussion forum activities.

Conclusion

Learner engagement can be an important indicator for academic performance and interest in online courses. The application of artificial intelligence to automatically assess learner engagement at scale lends valuable information on teaching and learning activities. Our study examines the changes in learners' social and cognitive presence in an asynchronous online discussion forum prior to and after the onset of COVID-19 lockdown. The emergent discussions around COVID-19 and quarantine experiences we detected in early spring quarter suggests discussion forum can serve as a rich source for understanding learner experience and emergent learning needs through online peer interactions. We demonstrate an analysis that combines multiple text mining techniques to effectively harvest

insights from learner discourse, sampling discussions surrounding specific topics for further qualitative analysis. We suggest the combination of latent semantic content and linguistic characteristics provides richer contextual details to learner interactions. This analytic process can be adapted to explore other research questions with different kinds of textual data. For instance, one may examine the impact of a change of instructional strategy or course requirement in using LMS. We may apply this analysis procedure to examine effects of such strategy on learners' online engagement before and after the change takes place. We may also analyze discourse characteristics and further connect them to learning outcomes and psychological wellbeing. In future studies, we plan to expand this analysis onto the entire discussion forum across disciplines, and examine the link between discourse features to survey response, so as to verify whether social, cognitive, and affective language predict learners' psychological experience (e.g. stress, perceived support) during COVID-19. The same analysis can also be applied to courses offered exclusively online where the level of asynchronous discussion might be higher, to see whether similar patterns can be found in learner discourse. Lastly, our study suggest that discussion forum holds promise for promoting genuine personal connection and building classroom community. We found that students engage in non-trivial discussions of COVID-19 in either a socially-oriented or cognitively-oriented manner. These discussion indicated a need for social connection and sense making, and emerged organically or facilitated by instructors. This suggest that teaching practices should be quickly adapted to meet future learners needs, and instructors should pay close attention to fostering meaningful learning experience. As more and more university courses remain online or hybrid to provide flexibility to students, instructors may consider facilitating meaningful social interaction by incorporating shared lived experiences or critical reflections on contemporary events, rather than superficial shallow connections for participation points in online discussion forums.

Funding This work was supported by The Andrew W. Mellon Foundation (1806-05902). The authors would like to thank the Next Generation Undergraduate Success Measurement Project team members for help with data collection.

Declarations

Conflict of Interest No conflict of interest is involved.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abrami, P. C., Bernard, R. M., Bures, E. M., Borokhovski, E., & Tamim, R. M. (2011). Interaction in distance education and online learning: Using evidence and theory to improve practice. *Journal Computing in Higher Education*, 23, 82–103. <https://doi.org/10.1007/s12528-011-9043-x>
- Adedoyin, O. B., & Soykan, E. (2023). Covid-19 pandemic and online learning: the challenges and opportunities. *Interactive learning environments*, 31(2), 863–875.
- Altrabsheh, N., Gaber, M. M., & Cocea, M. (2013). SA-E: sentiment analysis for education. *Intelligent Decision Technologies: Proceedings of the 5th KES International Conference on Intelligent Decision Technologies (KES-IDT 2013)*, (pp. 353–362). IOS Press.
- André, M., Mello, R. F., Nascimento, A., Lins, R. D., & Gašević, D. (2021). Toward automatic classification of online discussion messages for Social Presence. *IEEE Transactions on Learning Technologies*, 14(6), 802–816.
- Andrews, T. (2012). What is social constructionism? *Grounded Theory Review*, 11(1).
- Angelov, D. (2020). Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*. <https://doi.org/10.48550/arXiv.2008.09470>
- Ashokkumar, A., & Pennebaker, J. W. (2021). Social media conversations reveal large psychological shifts caused by COVID-19's onset across US cities. *Science Advances*, 7(39), eabg7843.
- Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, 79(3), 1243–1289.
- Bransford, J., Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. National Academies.
- Cazden, C. B. (1988). *The language of teaching and learning*. The language of teaching and learning, 2.
- Chong, C., Sheikh, U. U., Samah, N. A., & Sha'ameri, A. Z. (2020). Analysis on reflective writing using natural language processing and sentiment analysis. In *IOP Conference Series: Materials Science and Engineering* (Vol. 884, No. 1, p. 012069). IOP Publishing.
- Cohen, J. (1992). Quantitative methods in psychology: A power primer. In *Psychological bulletin*.
- Chopra, H. et al. (2022). Modeling student discourse in online discussion forums using semantic similarity based topic chains. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, V. Dimitrova (Eds.), *Artificial intelligence in education. posters and late breaking results, workshops and tutorials, industry and innovation tracks, practitioners' and doctoral consortium. AIED 2022. Lecture Notes in Computer Science*, vol 13356. Springer, Cham. https://doi.org/10.1007/978-3-031-11647-6_91
- Chong, C., Sheikh, U. U., Samah, N. A., & Sha'ameri, A. Z. (2020). Analysis on reflective writing using natural language processing and sentiment analysis. In *IOP Conference Series: Materials Science and Engineering* (Vol. 884, No. 1, p. 012069). IOP Publishing. <https://doi.org/10.1088/1757-899X/884/1/012069>
- Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Bio Medica: Atenei Parmensis*, 91(1), 157.
- Desai, V. P., Oza, K. S., & Kamat, R. K. (2021). Preference based e-learning during covid-19 lockdown: an exploration. *The online journal of distance education and e-learning*, 9(2), 285–292.
- Dolianiti, F.S., Iakovakis, D., Dias, S.B., Hadjileontiadou, S., Diniz, J.A., Hadjileontiadis, L. (2019a). Sentiment Analysis Techniques and Applications in Education: A Survey. In M. A. Tsitouridou, J. Diniz, T. Mikropoulos (Eds.), *Technology and Innovation in Learning, Teaching and Education. TECH-EDU 2018. Communications in Computer and Information Science* (vol 993). Springer. https://doi.org/10.1007/978-3-030-20954-4_31
- Dolianiti, F. S., Iakovakis, D., Dias, S. B., Hadjileontiadou, S. J., Diniz, J. A., Natsiou, G., ... & Hadjileontiadis, L. J. (2019b). Sentiment analysis on educational datasets: a comparative evaluation of commercial tools. *Educational Journal of the University of Patras UNESCO Chair*.
- Dowell, N., & Kovanovic, V. (2022). Modeling educational discourse with natural language processing. *Education*, 64, 82.
- Elmer, T., Mepham, K., & Stadtfeld, C. (2020). Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. *PLOS ONE*, 15(7), e0236337. <https://doi.org/10.1371/journal.pone.0236337>
- Ferreira, M., Rolim, V., Mello, R. F., Lins, R. D., Chen, G., & Gašević, D. (2020). Towards automatic content analysis of social presence in transcripts of online discussions. In *Proceedings of the Tenth*

- International Conference on Learning Analytics & Knowledge (LAK '20). Association for Computing Machinery* (pp. 141–150). <https://doi.org/10.1145/3375462.3375495>
- Garrison, D. R. (2009). Communities of inquiry in online learning. In P. Rogers, G. Berg, J. Boettcher, C. Howard, L. Justice, & K. Schenk (Eds.), *Encyclopedia of Distance Learning* (2nd ed., pp. 352–355). IGI Global. <https://doi.org/10.4018/978-1-60566-198-8.ch052>
- Garrison, D. R., & Arbaugh, J. B. (2007). Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and Higher Education*, 10(3), 157–172.
- Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *International Journal of Phytoremediation*, 21(1), 7–23.
- Gilbert, P. K., & Dabbagh, N. (2005). How to structure online discussions for meaningful discourse: A case study. *British Journal of Educational Technology*, 36(1), 5–18.
- Gkонтзис, А. F., Karachristos, C. V., Panagiotakopoulos, C. T., Stavropoulos, E. C., & Verykios, V. S. (2017). Sentiment analysis to track emotion and polarity in student fora. In Proceedings of the 21st Pan-Hellenic Conference on Informatics (PCI '17). Association for Computing Machinery, New York, NY, USA, Article 39, 1–6. <https://doi.org/10.1145/3139367.3139389>
- Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education*, 11(3), 8–26.
- Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education*, 145, 103724.
- Hilmy, S., De Silva, T., Pathirana, S., Kodagoda, N., & Suriyawansa, K. (2019). MOOCs recommender based on user preference, learning styles and forum activity. *2019 International Conference on Advancements in Computing (ICAC)* (pp. 180–185), Malabe, Sri Lanka. <https://doi.org/10.1109/ICAC49085.2019.9103376>
- Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M. S., Küfner, A. C., Deters, F. G., & Mehl, M. R. (2019). Linguistic markers of grandiose narcissism: A LIWC analysis of 15 samples. *Journal of Language and Social Psychology*, 38(5–6), 773–786.
- Hostetter, C. (2013). Community matters: Social presence and learning outcomes. *Journal of the Scholarship of Teaching and Learning*, 13(1), 77–86.
- Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216–225).
- Joksimović, S., Gasević, D., Kovanović, V., Adesope, O., & Hatala, M. (2014). Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *The Internet and Higher Education*, 22, 1–10.
- Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning*, 31(6), 638–654.
- Kang, M., Liew, B. T., Kim, J., & Park, Y. (2014). Learning presence as a predictor of achievement and satisfaction in online learning environments. In *International Journal on E-Learning* (Vol. 13, No. 2, pp. 193–208). Association for the Advancement of Computing in Education (AACE).
- Kilis, S., & Yıldırım, Z. (2018). Investigation of community of inquiry framework in regard to self-regulation, metacognition and motivation. *Computers & Education*, 126, 53–64.
- Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16). Association for Computing Machinery*, New York, NY, USA, pp. 15–24. <https://doi.org/10.1145/2883851.2883950>
- Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning*, in *Proceedings of Machine Learning Research* 32(2), 1188–1196. Available from: <https://proceedings.mlr.press/v32/le14.html>
- Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, 2, 100016.
- Lin, Y., Yu, R., & Dowell, N. (2020). LIWCS the same, not the same: Gendered linguistic signals of performance and experience in online STEM courses. In I. Bittencourt, M. Cukurova, K. Mulder, R. Luckin, & E. Millán (Eds.), AIED 2020. *Lecture Notes in Computer Science()*, Vol. 12163. Springer, Cham. https://doi.org/10.1007/978-3-030-52237-7_27

- Lundqvist, K., Liyanagunawardena, T., & Starkey, L. (2020). Evaluation of student feedback within a MOOC using sentiment analysis and target groups. *International Review of Research in Open and Distributed Learning*, 21(3), 140–156.
- Markowitz, D. M. (2023). Instrumental goal activation increases online petition support across languages. *Journal of Personality and Social Psychology*, 124(6), 1133.
- Mishra, L., Gupta, T., & Shree, A. (2020). Online teaching-learning in higher education during lockdown period of COVID-19 pandemic. *International Journal of Educational Research Open*, 1, 100012.
- Moore, R. L., Oliver, K. M., & Wang, C. (2019). Setting the pace: Examining cognitive processing in MOOC discussion forums with automatic text analysis. *Interact Learn Environ*, 27(5–6), 655–669.
- Monzani, L., Escartín, J., Ceja, L., & Bakker, A. B. (2021). Blending mindfulness practices and character strengths increases employee well-being: A second-order meta-analysis and a follow-up field experiment. *Human Resource Management Journal*, 31(4), 1025–1062.
- Monzani, A., Ragazzoni, L., Della Corte, F., Rabbone, I., & Franc, J. M. (2020). COVID-19 pandemic: Perspective from Italian pediatric emergency physicians. *Disaster medicine and public health preparedness*, 14(5), 648–651.
- 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), 8438.
- Munezero, M., Montero, C. S., Mozgovoy, M., & Sutinen, E. (2013). Exploiting sentiment analysis to track emotions in students' learning diaries. In Proceedings of the 13th Koli Calling International Conference on Computing Education Research (Koli Calling '13). Association for Computing Machinery (pp. 145–152). <https://doi.org/10.1145/2526968.2526984>
- Mustadi, A., Annisa, F. C., & Mursidi, A. P. (2021). Blended learning innovation of social media based active English during the COVID-19 pandemic. *Ilkogretim Online*, 20(2), 74–88.
- Neumann, M., & Linzmayer, R. (2021). Capturing student feedback and emotions in large computing courses: A sentiment analysis approach. In Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21). Association for Computing Machinery (pp. 541–547). <https://doi.org/10.1145/3408877.3432403>
- Ouyang, F., & Chang, Y. H. (2019). The relationships between social participatory roles and cognitive engagement levels in online discussions. *British Journal of Educational Technology* 50(3), 1396–1414.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1), 547–577.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, University of Texas at Austin.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS one*, 9(12), e115844.
- Peng, X., & Xu, Q. (2020). Investigating learners' behaviors and discourse content in MOOC course reviews. *Computers & Education*, 143, 103673.
- Pokhrel, S., & Chhetri, R. (2021). A literature review on impact of COVID-19 pandemic on teaching and learning. *Higher Education for the Future*, 8(1), 133–141.
- Räaf, S. A., Knöös, J., Dalipi, F., & Kastrati, Z. (2021). Investigating learning experience of MOOCs learners using topic modeling and sentiment analysis. In *2021 19th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 01–07). IEEE. <https://doi.org/10.1109/ITHET50392.2021.9759714>
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27.
- Romero, C., & Ventura, S. (2017). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(1), e1187.
- Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *The Journal of Distance Education/Revue De L'éducation Distance*, 14(2), 50–71.
- Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 26.
- Seraj, S., Blackburn, K. G., & Pennebaker, J. W. (2021). Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences*, 118(7), e2017154118.

- Song, D., Lin, H., & Yang, Z. (2007). "Opinion mining in e-learning system". In *2007 IFIP International Conference on Network and Parallel Computing Workshops* (NPC 2007) (pp. 788–792). <https://doi.org/10.1109/NPC.2007.51>
- Strielkowski, W. (2020). *COVID-19 pandemic and the digital revolution in academia and higher education I*, 1–6. Preprints.
- Sun, Y., Lin, S. Y., & Chung, K. K. H. (2020). University Students' perceived peer support and experienced depressive symptoms during the COVID-19 pandemic: The mediating role of emotional well-being. *International Journal of Environmental Research and Public Health*, 17(24), 9308.
- Suryaningsih, V. (2021). Strengthening student engagement: How student hone their soft skill along online learning during Covid-19 pandemic? *Jurnal Manajemen Bisnis*, 18(1), 1–15.
- Swan, K. (2010). Post-industrial distance education. In R. Garrison & M. Cleveland-Innes (Eds.), *An introduction to distance education: Understanding teaching and learning in a new era* (pp. 108–134). Routledge.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, 23(3), 34–41.
- Windsor, L. C., Cupit, J. G., & Windsor, A. J. (2019). Automated content analysis across six languages. *PloS One*, 14(11), e0224425.
- Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using latent dirichlet allocation for topic modeling on Twitter. *PloS one*, 15(9), e0239441.
- Ziegler, M. F., Paulus, T., & Woodside, M. (2014). Understanding informal group learning in online communities through discourse analysis. *Adult Education Quarterly*, 64(1), 60–78.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.