Article

Sentiment Analysis and Topic Modeling Regarding Online Classes on the Reddit Platform: Educators versus Learners

Shanghao Li [1], **Zerong Xie** [1], **Dickson K. W. Chiu** [1] **and Kevin K. W. Ho** [2,*]

[1] Faculty of Education, The University of Hong Kong, Hong Kong 999077, China
[2] Graduate School of Business Sciences, Humanities and Social Sciences, University of Tsukuba, Tokyo 112-0012, Japan
* Correspondence: ho.kevin.ge@u.tsukuba.ac.jp

Abstract: The world is witnessing an unpredictable COVID-19 pandemic that has impacted all levels of online education, shaping future trends. However, this shift was so sudden and drastic that unrevealed puzzles exist regarding the public's authentic opinion towards online classes, even though three years have passed. Many experts and policymakers have conducted qualitative and quantitative research to explore effective pedagogies, the satisfaction of different stakeholders, and factors influential on learners' performance. However, scant studies have examined personal opinions and concerns toward online classes hidden behind people's anonymous posts on social media. This research investigates the sentiments, concerns, and their variance with time regarding online classes by learners and educators on Reddit, which is a dominant social network among them. Data were collected via the official API from identified relevant subreddits and keyword search results across Reddit. Sentiment analysis was applied to reveal their emotions and their changes. Topic modeling was conducted to discover the concerns hidden in the posts. The results revealed the concerns about online classes, such as severe cheating behaviors, and showed doubts about previous strategies to solve disadvantages in online classes. In addition, the results verified the habitual difference and motivations of social media usage between educators and learners.

Keywords: online classes; sentiment analysis; topic modeling; data mining; content analysis; social media

1. Introduction

Due to the COVID-19 pandemic, most countries implemented lockdown and social distancing measures, resulting in the closure of schools, training institutions, and further educational activities [1–3]. The paradigm shift from traditional face-to-face to online classes poses various difficulties for educators and learners, leading to various pedagogical innovations supporting continuing online classes [4]. Yet, the reliability and efficiency of pedagogy for online classes heavily depend on educators' and learners' utilization and exposure to information and communications technology (ICT) [5,6]. Microsoft Teams, Google Classroom, Zoom, Canvas, and Blackboard are widely-used communication and collaboration platforms to facilitate video meetings, file sharing, storage, quizzes, and rubric-based assessments [7,8]. Additionally, the flipped classroom is widely adopted for learners to review learning resources such as articles, pre-recorded videos, and YouTube tutorials before classes and deepen their understanding through discussion with peers and educators during online classes [9,10].

However, there are concerns about the effectiveness and adaptability of online classes [6]. For example, educators struggle with conducting online classes due to a lack of proper training and infrastructure [11]. While open-minded learners quickly adapt to a new learning environment, those with fixed mindsets encounter difficulties adjusting and adapting. Without a widely adopted pedagogy for online classes, the readiness of various stakeholders needs to be flexible and supported accordingly [5]. Research investigating

factors impacting online classes indicates that motivation is the most significant factor, contributing to an almost 50% impact on learner performance in online classes. Thus, the biggest challenge for educators is motivating learners to focus in classes [12]. Many learners prefer conventional to online classes because they face challenges in getting a stable Internet connection, understanding the teaching content, interacting with classmates and educators [13], and building up their social networks for social support [14]. Many even believe online classes cannot completely replace traditional schooling [6,11,15].

With the diverse views and stakeholders, it is necessary to investigate broad opinions beyond empirical studies on social media, and similar studies are lacking in this area. Thus, this study employs a social media analytical approach [16–18] to investigate the sentiments, concerns, and their variance with time regarding online classes by learners and educators.

2. Literature Review

Reddit is among the most popular social networks, with over 50 million daily active users and more than 100,000 active topical communities called "subreddits." Its image is associated with the people's forum, a warehouse for the dankest memes, the weapon against Wall Street, and the "the front page of the Internet" [19]. Being valued at up to USD 15 billion, Reddit is on the way to going public (possibly as soon as this year), with preliminary IPO registration statements already filed under the guidance of Morgan Stanley and Goldman Sachs [20].

Although the demographic features of the 330 million active users are unavailable because of its policy of not collecting personal information, evidence shows that a majority of users are young people between 18 and 34 years old (57%) [21], corresponding to the age of college learners. In addition, research on educators on Reddit reveals that educators widely use the site and that its use might be influential in training preservice educators [22].

Given Reddit's accessibility and popularity and the capability to collect high-quality data [23], a growing volume of research has used Reddit as a data source in the past decade. These studies used different types of data, including the original content, comments and comment threads, meta-data, links or media, upvoting/downvoting information, characteristics of subreddits, as well as surveys with users and moderators [21]. However, few researchers share the datasets drawn from Reddit. Notably, Proferes et al. [21] provide procedure descriptions for gathering their data, the metadata of the final dataset, and even the list of subreddits selected as the sources.

The subreddits of interest vary with the specific research context, with political subreddits, mental health subreddits, and drug use subreddits being more prominent data sources [23]. However, Reddit has evolved into a widespread forum with diverse topics, including education [24] and much research in computer science, engineering, and mathematics disciplines with computational-driven textual analysis [21].

Discussions on Reddit are open to anyone with or without a Reddit account unless the subreddits' setting is private. The visibility of both the original post and discussion comments is configurable by users' "upvote" and "downvote" behaviors. Registration for a Reddit account only requires a unique username and a password without email authentication [21]. Users are allowed to post content anonymously under a pseudonymous account or multiple accounts [25], facilitated by "throwaway accounts" when the topic is particularly sensitive or personal, encouraging individuals to engage in more sensitive, potentially stigmatizing conversations [26].

3. Research Gap

On the other hand, studies on online classes during COVID-19 mainly focus on innovative pedagogies and examination effectiveness. Few investigate satisfaction and concerns from different stakeholders. A few studies examined some aspects of education-related issues on forums to help better understand various stakeholders' opinions, such as COVID-19 vaccination-related issues [27]. However, to our knowledge, no study examines online educational issues in the forum context. Even worse is that most studies adopted

traditional approaches to collect empirical data, such as surveys and interviews, which may contain response bias and limited generalizability. Given that educators and learners spend more time online, more effort is necessary to gather online self-disclosure posts for further analysis.

Reddit is one of the most popular social media among younger generations and education practitioners. Its unique anonymity mechanism gives users enough safety to express their genuine feelings and opinions. However, previous studies concentrate on the technology field, undertreating the potential of Reddit as a data source in the education discipline. As such, there is a need to comprehend the authentic sentiments among different stakeholders in online classes and the obstacles to conducting effective online classes through Reddit. Three research questions were defined to guide this research and reach the research goals as listed below:

RQ1: What topics can be detected from Reddit posts about online classes?
RQ2: What are the differences in sentiments and topics between different stakeholders?
RQ3: How do the sentiments and topics vary with time?

4. Methodology

4.1. Data Collection

There are two ways to request data from Reddit: pre-stored third-party data collection Pushshift (https://github.com/pushshift; accessed on 12 May 2022) and official Reddit API. Pushshift is a social media archiving tool that can gather Reddit data and make it accessible to researchers. Data copied into Pushshift when submitted to Reddit is updated in real-time and complete [28]. Everyone can access Reddit data through Pushshift API or the downloadable dataset without restrictions. The Reddit office powers Reddit API to encourage the developer community to participate in building great products. It enables programmatic control of nearly every function users can perform on the site. This API requires an authorization application that is easier, separated into several steps, and without a long-time review process. Table 1 summarizes the differences between the two approaches.

Table 1. Differences Between Pushshift.io and Reddit API.

Feature	Pushshift.io	Reddit API
Data Scope	• All content available on Reddit, including posts that may have been removed or modified (not available on the site now)	• Only the content currently available on the site
Accessibility	• Pushshift API • Downloadable dataset • PSAW, a Python Pushshift API wrapper	• Reddit API • PRAW, a Python Reddit API wrapper
Data Quality	• The downloadable dataset aligns with the FAIR principles. • Meta-data of the submissions (e.g., score and awards) might be outdated [29]. • The data structure and contents are not consistent over time owing to the changes in Reddit itself.	• The data is more reliable and more regularly updated since it is derived from the official storage. • Historical data between specific dates is not available.
Restriction	• Max pull request at 60 items/min • No return limitations. All the data stored in the database is retrievable	• Max pull request at 60 items/min • Return at most 1,000 items

In summary, Reddit API is easier to use and more regularly updated, but it only allows the extraction of data from up to 1000 posts at a time. Using Pushshift API is slightly more

cumbersome, but a larger amount of historical data is available. Since this research covered data over two years, Pushshift was selected for data collection.

4.2. Data Source Selection

The initial examination of subreddits relevant to the extensive topics about online classes involved a subreddit search using the official search API with the term **online class* in which '*' serves as the wildcard character. Fifty-seven subreddits were identified as associated with online classes. However, some results were not useful for this research, e.g., *r/softwaregore* is for sharing funny software vulnerabilities, and *r/memes* content is in the form of images, both infeasible for textual analysis. Note that on the Reddit website, the symbol '*r/*' denotes a subreddit of the name following it. For example, *r/softwaregore* is a subreddit called *softwaregore*.

Next, each identified subreddit was further assessed for its appropriateness for further mining and analysis based on: (1) whether subscribers are representative of different stakeholders; (2) whether different users posted a large amount of content instead of bots or specific user groups; and (3) the average quality of content measured by the length of text, the number of comments, and the upvote and downvote score. As a result, six subreddits were finally chosen as the data sources, as detailed in Table 2.

Table 2. Subreddits chosen for analyses.

Subreddit	Populations	Creation Time	Subscribers	Descriptions
r/funnyonlineclasses	Learners, Educators, Parents	28 March 2020	14.8k	Anything funny about online classes in the wake of the COVID-19 quarantine.
r/teenagers	Learners	27 February 2010	30.8k	Biggest community forum run by teenagers for teenagers.
r/college	Learners	25 January 2008	806k	For discussion related to college and collegiate life.
r/CollegeRant	Learners	7 August 2018	32.2k	For sharing experiences in college to discuss the negative aspects of college life.
r/Teachers	Educators	23 December 2008	333k	A sub for all things educator related.
r/Professors	Educators	14 September 2011	101k	BY professors FOR professors.

Additionally, threads in terms of online classes in a global context were also collected to reflect public perspectives using the post-search feature of Pushshift API. The period covered 1 January 2020 to 30 April 2022.

4.3. Data Analysis

We analyzed our data using three processes, i.e., data preprocessing, sentiment analysis, and topic modeling.

Data preprocessing was performed with the Regex library (https://regexlib.com; accessed on 12 May 2022) to tokenize, clean, and format the text of each subset. The first step was to tokenize the data, removing special characters and hyperlinks and segmenting sentences into individual words. Then, all stopwords, i.e., commonly used non-emotional words (e.g., I, a, the), were removed. We adapted NLTK's *stopwords* corpus to identify and filter the stopwords. Next, the remaining tokens were standardized, with all capital letters converted into lowercase, contractions reverted to their composite words (e.g., don't to do not), and numbers replaced by their corresponding word (e.g., 1 to one). We adapted NLTK's *tokenize* package (https://www.nltk.org/api/nltk.tokenize.html; accessed on 12 May 2022) to implement this standardized process. The final step, called lemmatization, initialized standardized tokens into their root word form (e.g., taught to teach).

Sentiment analysis is an approach for automatically extracting and classifying sentiment from textual documents using natural language processing (NLP), textual analysis, and computational techniques [30]. This approach can avoid taking keywords out of context and consider users' feelings at the post level. This study selected VADER (Valence

Aware Dictionary and sEntiment Reasoner) because it is specifically tuned to the social media context and supports web expressions such as emoticons and shorthand [31]. The result returned by VADER was a dictionary of four keys: *neg*, *neu*, *pos*, and *compound*. *Compound* is computed by normalizing the other three scores. By adjusting the threshold based on the dataset to the compound value, the sentiment of a complete post could be split into negative, neutral, and positive.

To evaluate the effectiveness of the application of VADER to our dataset, we sampled 100 pieces of data from the global data set and labeled it manually into three classes, i.e., "negative," 'neutral,' and 'positive,' and calculated the prediction accuracy. We used accuracy as the metric because the classes of our data set were relatively balanced in terms of label distributions. In addition, there was no difference between the classes for decision-making implications in our research context, as we only used the classification results to plot the sentiment-time line chart. Therefore, we did not calculate the terms of precision and recall. We used the following equation to calculate the accuracy.

$$Accuracy = (cp + 0.5ipn)/total\ number\ of\ samples \tag{1}$$

The *cp* in the above equation (1) indicates the correct number of predictions, and *ipn* indicates the incorrect number of predictions involving the 'neutral' manual labels. The reason to add the *0.5ipn* part to calculate the accuracy is to cater to the ranking relations in the three classes, i.e., 'negative,' 'neutral,' and 'positive.' Therefore, it will be fairer to assign half of the marks to the failed 'neutral' cases, compared to giving the same zero mark to the failed cases predicting "negative" while the true value is "positive" and vice versa.

Topic modeling is a statistical approach for discovering "topics" hidden in the text corpus, which is the cleaned dataset in this research. Latent Dirichlet Allocation (LDA) is one of the most popular topic models, which creates two Dirichlet distribution models; one represents the topic in the document, and the other represents the words in each topic. Next, the Python machine-learning module *scikit-learn* (https://scikit-learn.org; accessed on 12 May 2022) was imported to create the LDA topic model. First, DTM (Document-Term Matrix) was calculated to indicate the frequency that a term occurs in the document. Then, TF-IDF (Term Frequency-Inverse Document Frequency) was computed to evaluate the term's relevance to a certain document by two metrics: how many times a term appeared in a certain document (TF) and how many times it appeared across the entire document set (IDF). The final LDA topic model was generated by populating the DTM with the TF-IDF score and optimizing the result by modifying parameters. To visualize the optimal result, pyLDAvis (https://github.com/bmabey/pyLDAvis; accessed on 11 June 2022) was applied to create an interactive figure in which the top 30 most relevant terms were displayed according to frequency once a topic was selected, in addition to static charts and word clouds drawn by the pandas and seaborn libraries.

5. Results

5.1. Data Collected

After the collecting and filtering process, 19,818 posts posted from 1 January 2020 to 30 April 2022 with title and body text finally remained (retrievable at https://github.com/hellotum/learnerEducatorData, with the data set description in the file README.md). Table 3 illustrates the number of posts collected from the six subreddits and global. Additional attributes of posts include post time, number of comments, and author. Nine datasets were built and saved into CSV files, among which seven were directly built upon data collected from global and subreddits, and two were further integrated to represent different stakeholders. The learner dataset was created by combing the data from three subreddits, including *r/teenagers*, *r/college*, and *r/CollegeRant*, while the educator dataset was from *r/Teachers* and *r/Professors*. Table 4 shows the first five rows of the global data set as the sample of the posts.

Table 3. Number of posts from different sources.

Source	Posts Count
global	6162
r/funnyonlineclasses	375
r/teenagers	5620
r/college	4077
r/CollegeRant	690
r/Teachers	1800
r/Professors	1094
Total	**19818**

Table 4. Sample posts.

Time	Subreddit	Title	Body	Number of Comments	Score
29 April 2022 11:25:30	r/studentsph	Ako lang ba ang parang "nabobo" dito sa online . . .	Like tangina, feel ko napaka walang kwenta net . . .	0	1
29 April 2022 09:29:20	r/AskReddit	Which games did you play during online class?		0	1
29 April 2022 08:42:32	r/Philippines	Any amount will help for my online class Thank . . .		0	1
29 April 2022 08:37:08	r/Philippines	any amount will help for my online class		0	1
26 April 2022 18:17:13	r/studentsph	What are good excuses for being absent from an . . .	I think I'm running out of believeable excuses, . . .	0	1

5.2. Topics Detected (RQ1)

By observing the terms associated with each topic and the inter-topic distance when assigning different topic numbers, batch sizes, and random states, six topics are finally determined as the optimal result. Figure 1 shows three of the six topics, Topics 3, 5, and 6, distinct from each other, while there is an overlap between the three remaining topics (Topics 1, 2, and 4). Figure 2 displays the fifteen most relevant terms for each topic. Figure 3 depicts the Word clouds generated from the six topics.

Among the three distinctive topics, Topic 3 is associated with *teacher*, *thing*, *happen*, and various adjective terms such as *funniest*, *weirdest*, and *embarrassing*. It reflects the abnormal, strong, and diverse incidents and interactions between learners and educators during online classes. Topic 5 is featured with the terms *help*, *exam*, and *post*, which implies that learners seek help in exams and assignments when taking online classes. The salient terms *math*, *statistic*, and *software* indicate that learners experience more difficulties in science and engineering classes, with *meme*, *made*, *time*, and *today* being the most dominant terms in Topic 6. It illustrates that online class has become a popular resource for meme-making. In other words, it has become a part of the subculture among the young generation.

Among the three overlapping topics, Topic 1 covers the largest number of words, featuring *teacher*, *student*, and *time*, about small things frequently happening in online classes. Topics 2 and 4 are closely associated with each other. Topic 4 includes the *bored*, *tifu*, and *camera*, which emphasize the negative aspect of daily online classes, while Topic 2 is more positive since it includes *like*, *friend*, *love*, and *fun*.
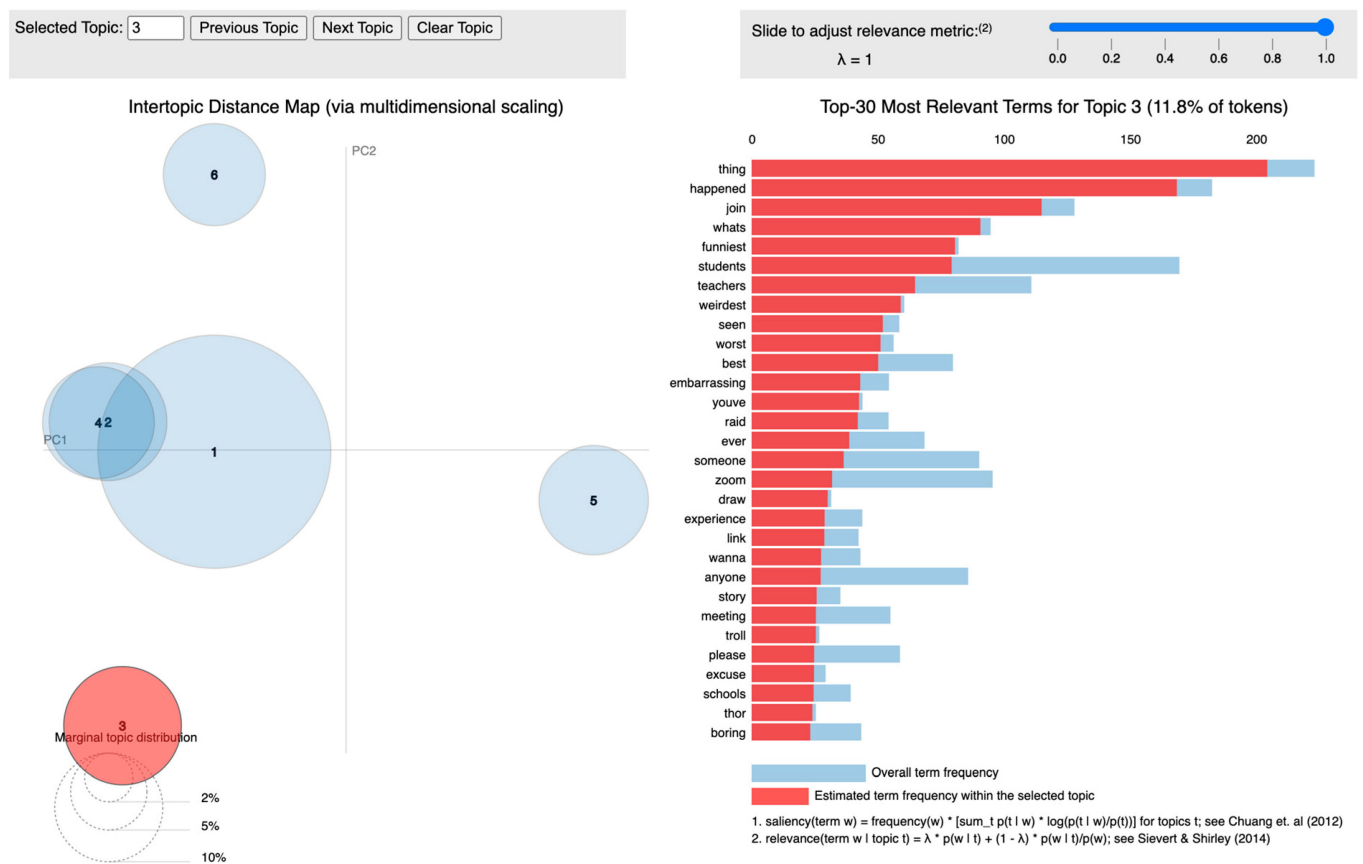
Figure 1. Interactive topic visualization created by pyLDAvis.

	Topic 1	Weights 1	Topic 2	Weights 2	Topic 3	Weights 3	Topic 4	Weights 4	Topic 5	Weights 5	Topic 6	Weights 6
0	classes	112.7	like	50.5	thing	133.2	bored	76.7	math	46.5	made	228.8
1	like	74.9	teacher	33.4	happened	109.9	teacher	59.6	moment	33.2	meme	66.4
2	know	64.4	friend	32.8	join	74.9	got	43.8	help	28.3	found	29.4
3	time	62.6	middle	22.4	whats	59.1	tifu	36.6	exam	24.0	today	27.1
4	one	58.3	minutes	22.0	funniest	52.6	camera	30.1	weeks	23.4	memes	26.2
5	dont	57.7	boys	21.8	students	51.7	kid	29.6	homework	21.7	background	20.4
6	school	50.9	guy	21.3	teachers	42.2	posted	22.4	hate	19.9	attending	18.5
7	really	47.2	taking	20.0	weirdest	38.5	think	20.9	discussion	17.9	first	17.3
8	want	44.7	drew	19.2	seen	33.9	playing	18.8	quiz	17.0	rick	14.7
9	first	44.7	thought	17.4	worst	33.3	upvote	17.3	discord	17.0	dog	12.8
10	every	41.8	girl	16.3	best	32.7	lol	16.7	statistics	16.5	going	12.3
11	take	41.2	love	16.1	embarrassing	28.1	posting	16.7	algebra	15.2	ako	11.5
12	teacher	39.7	fun	15.9	youve	27.8	mic	16.3	software	15.2	rolled	11.3
13	people	38.8	got	14.7	raid	27.4	average	15.7	chat	14.3	bad	11.1
14	students	38.5	pls	14.5	ever	25.2	see	15.6	take	14.2	mga	10.9

Figure 2. Top 15 most salient terms of each topic.

Figure 3. Word cloud of each topic.

5.3. Sentiment Difference between Stakeholders (RQ2)

We conducted the sentiment analysis, which resulted in a 78% prediction accuracy with our samples. We then used the classification result for further analysis. Figure 4 shows the sentiment distributions of posts in the six subreddits. Interestingly, none of the six distributions coincide with the normal distribution. Instead, there are sharp sentiment polarizations in the five subreddits, except for *r/funnyonlineclasses*. Users seem to express intense and differentiated emotions towards online classes on Reddit. This result verified the previous finding that sharing beliefs and opinions to find others who share similar ideas is a significant motivation for Reddit users to post content [32]. Figures 5 and 6 display more detailed sentiment distributions of each subreddit, with a single post's sentiment value counted and indicated by a dot and bar.
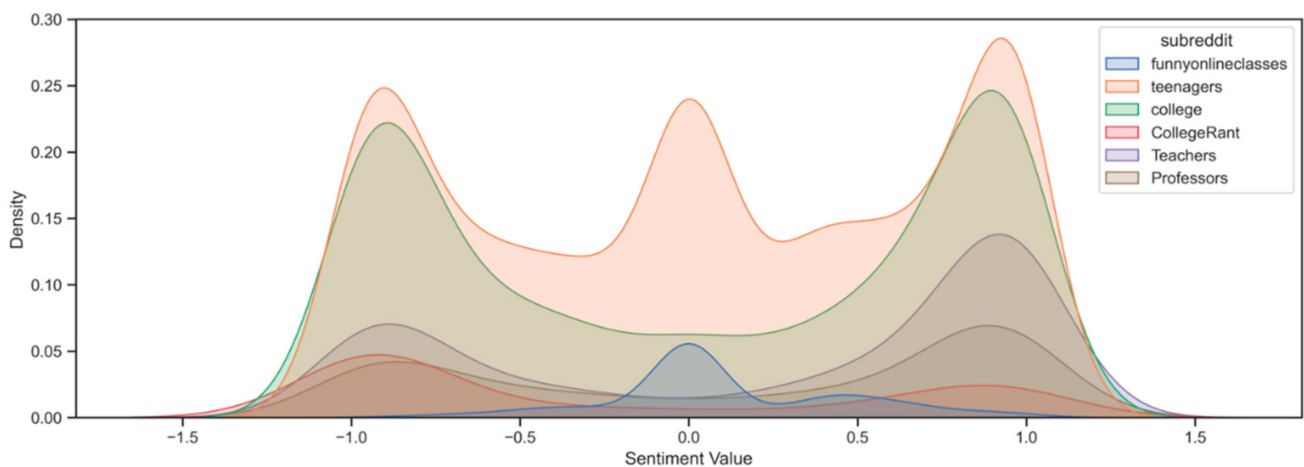

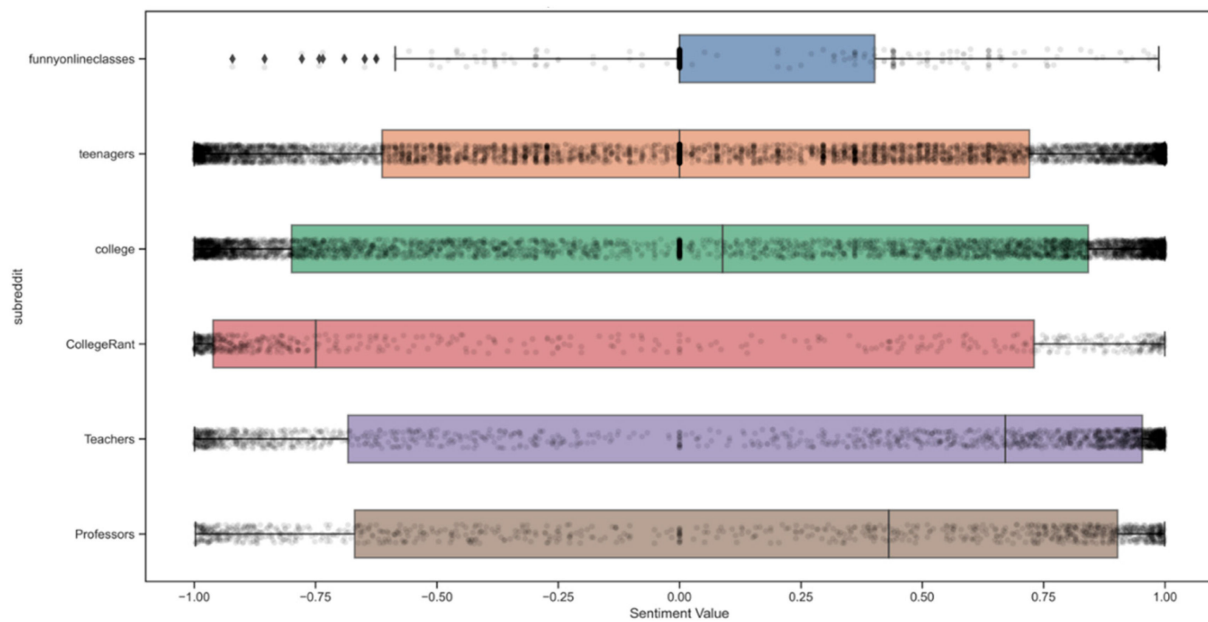**Figure 4.** Sentiment distributions of the six subreddits.

Figure 5. Box plot of posts' sentiment distributions of six subreddits.
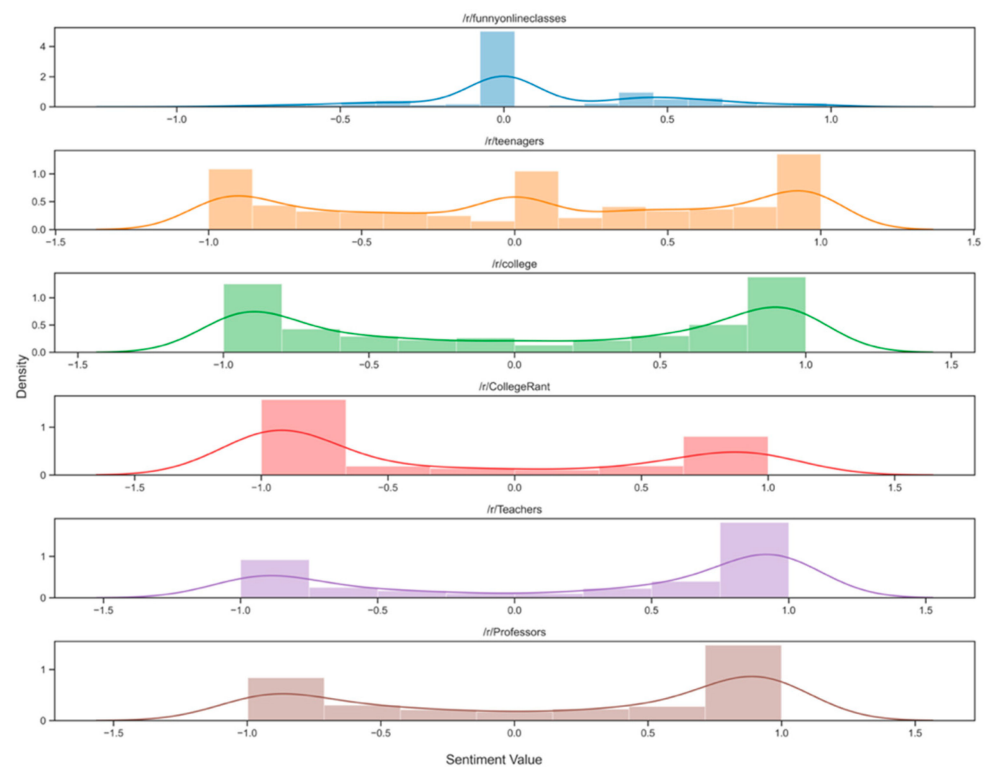

Figure 6. KDE plot of posts' sentiment distributions of six subreddits.

Compared with the other five subreddits, *r/funnyonlineclass* illustrates the most normalized distribution, where most posts display neutral and slightly positive sentiment, which appears reasonable. Surprisingly, among the three subreddits composed of learners of different ages, the sentiment of *r/teenagers* is the most gentle. The median sentiment value of this subreddit is neutral, even though a significant amount of sentiment is either very positive or negative. Yet, the other two subreddits, *r/college* and *r/CollegeRant*, show extreme polarization. Notably, *r/college* demonstrates a similar amount and extent of positive and negative emotions, whereas *r/CollegeRant* shows a higher number of negative emotions. This result can be explained by this subreddit's topic on the negative aspects of college life.

On the educators' side, positive sentiment exceeds negative sentiment. Moreover, with a higher median sentiment value and higher density of positive sentiment value, *r/Teachers* expresses stronger negative emotions than *r/Professors*.

After cleaning and tokenizing the sentence in the two representative datasets, the sentiment of different words in the posts was calculated based on negative, neutral, and positive word classification. The negative words and positive words in the two datasets were ranked according to the frequencies.

Further analysis was conducted on the learner and educator datasets to understand the two types of stakeholders. Figure 7 shows the kernel density estimate (KDE) plot of both sides for comparison. Overall, learners express more extreme sentiments, while educators tend to have more positive sentiments. The difference in motivations behind social media usage could explain this: educators generally use social media for self-development purposes, such as sharing resources and seeking collaborations with potential colleagues [22].
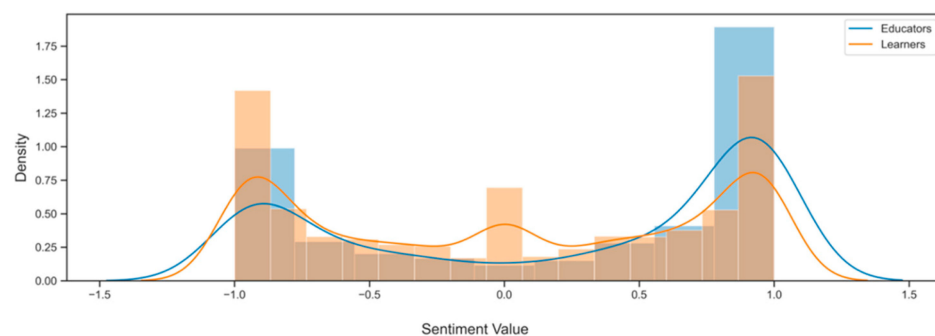

Figure 7. KDE plot of sentiment distributions of educators and learners.

Figure 8 shows the sentiment of words constituting posts extracted and ranked by frequency. Notably, educators and learners shared a similar set of sentimental words, such as *bad*, *hate*, *shit*, *fail*, and *hell*, widely used for expressing negative feelings, and *best*, *love*, *great*, *kind*, and *free* to express positive feelings in contrast. However, the term *cheat* is particularly concerning since it ranks high on both sides.

	educators negative words	frequency	learners negative words	frequency	educators positive words	frequency	learners positive words	frequency
0	bad	360	bad	1893	best	492	best	1434
1	failing	226	shit	1388	love	434	love	999
2	hate	218	hate	1247	great	390	great	889
3	sick	209	fuck	942	kind	309	kind	874
4	cheating	169	fail	701	free	232	fun	771
5	shit	161	stupid	701	positive	210	happy	648
6	fail	139	failed	699	fun	204	free	646
7	hell	137	failing	555	happy	163	super	551
8	failed	122	worst	467	super	149	appreciated	229
9	frustrated	110	hell	465	appreciated	140	perfect	225
10	worst	104	ass	442	amazing	87	honest	195
11	stupid	91	cheating	405	honest	75	amazing	192
12	fuck	79	depression	398	loved	70	positive	191
13	angry	78	sick	366	success	70	loved	187
14	negative	75	depressed	343	comfortable	67	honors	168
15	horrible	59	fucked	314	perfect	66	trust	155
16	panic	54	dumb	305	awesome	62	comfortable	149
17	ass	54	horrible	261	encourage	51	confidence	143
18	failure	53	angry	250	successful	45	ideal	140
19	bullshit	50	shitty	240	confidence	44	laugh	136

Figure 8. Word sentiment in educators and learners ranked by frequency.

Posts regarding cheating were extracted, and Appendix A shows several representative examples. Educators observed widespread cheating behavior and this should be taken seriously as a systemic problem. Accordingly, we propose various practical approaches to prohibit cheating and care for learners' mental health. On the learners' side, motivations for cheating are complicated. It may be due to their difficulties studying online and fear of exam failure. Furthermore, cheating is also a social behavior driven by others' influence because their classmates around them cheat. They fear that they may fall behind if they do not do so. Some are even under pressure to help others cheat. In addition to the identified topic of asking for help in exams and assignments, results indicated that cheating behavior is much more severe than expected.

Notably, results showed several unique sentimental words on both sides, e.g., *panic*, *successful*, and *awesome* for educators and *depression*, *depressed*, *dumb*, *trust*, and *laugh* for learners. These words may explain the nuanced difference in the sentiment between educators and learners. Posts containing the above unique keywords were extracted for further qualitative analysis, with some examples summarized in Appendix B.

The term "panic" reflects the following worries. Extra procedures of class preparations brought by software result in panic among many educators. Learners' panic and scarce participation even worsen this panic. The terms *successful* and *awesome* are associated with those who actively share effective experiences, give encouragement, and ask for suggestions. Among learners, the terms *depress* and *dumb* revealed their terrible mental state. They felt it was difficult to finish assignments and pass exams and blamed themselves for this. At the same time, a lack of communication among classmates made them feel lost and left behind. The positive term *trust* commonly came from supportive posts, which indicated hard work and self-confidence. Behind the term *laugh*, learners try to make the class enjoyable.

5.4. Changes over Time (RQ3)

Figure 9 illustrates the change in post numbers over time on Reddit global. The dramatic increase in online-class-related posts started at the beginning of March 2020 when the worldwide lockdown of educational institutions spread. After peaking at the end of that month, the number kept decreasing until it reached its lowest point in June 2020, when summer vacation was about to start. The subsequent two apparent fluctuations coincided with the arrangement of the second online semester. The first peak occurred at the end of August 2020, the start of a new semester, and the other peak in the middle of November 2020 was the time of final exams. The result triangulates our findings that learners struggle with online examinations. In 2021, with the highest number at the start of the first semester, the number gradually decreased as the lockdown was loosened worldwide. In 2022, the number remained at a much lower level. The general decreasing trend in the number of posts is probably due to the smaller amount of time for online classes instead of the improvement in class effectiveness.
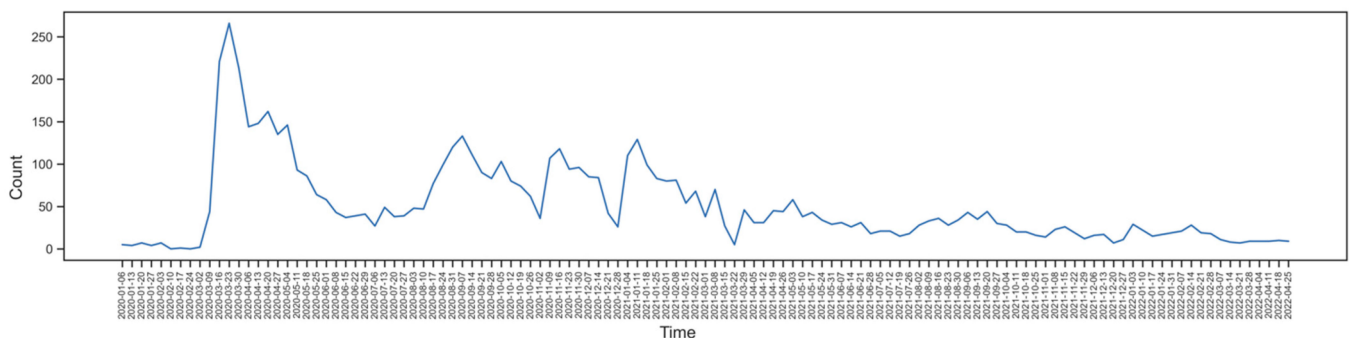

Figure 9. Number of posts varying with time in Reddit global.

However, Figure 10 shows the differences between educators and learners. In general, the features of the post number, including the peaks, bottoms, increases, and decreases in both parties, were similar to the global trend. Nevertheless, the learner's trend was much closer to the global trend, with a much lower number of posts posted in the latter two years. In contrast, the post number of educators was maintained at a certain level, without any downward trend, indicating online class is still a focus of attention among educators, though not prevalent as before.

Figure 10. Number of posts for educators and learners.

Figures 11 and 12 show the changes in global post numbers by sentiment type. Combined with Figure 8, neutral posts constitute the majority of posts, thereby decreasing the entire post number. Except for the fluctuation at the initial stage, there is no strong fluctuation in the number of negative and positive posts. Overall, the number of neutral posts decreases as time goes by, with fewer online classes, whereas the numbers of negative and positive posts are relatively stable. This observation further validates the analysis that the post number is correlated with the popularity of online classes and indicates that there are still unsolved concerns about online classes.
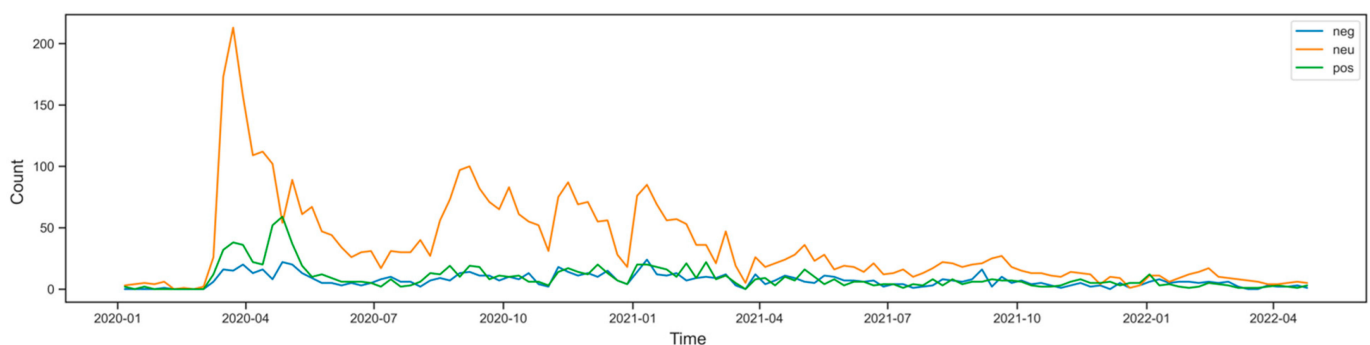

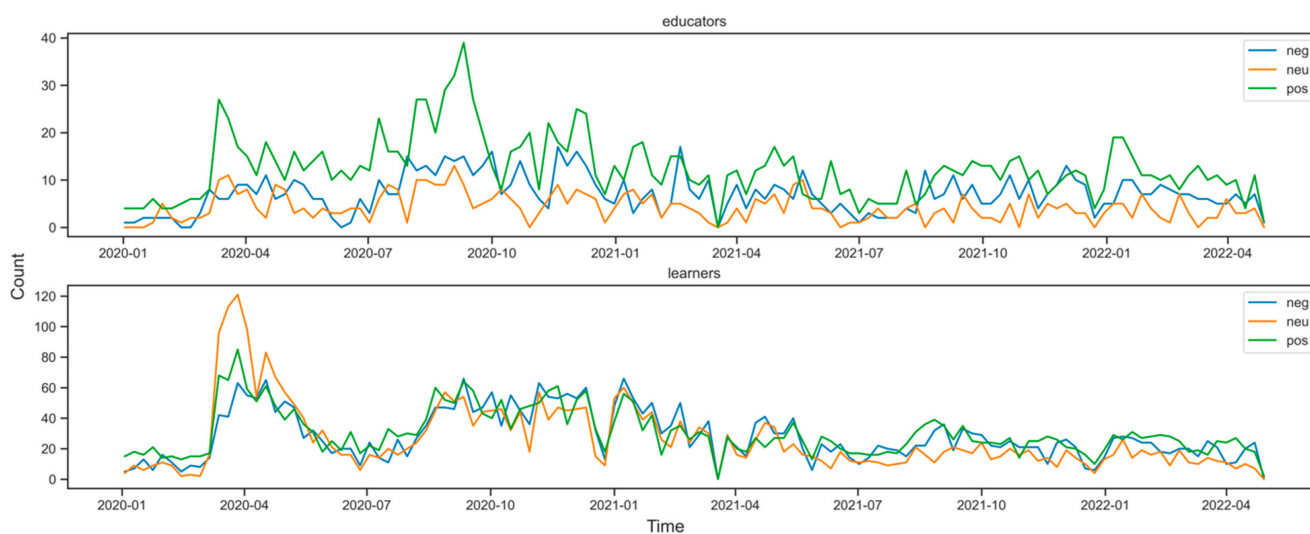**Figure 11.** Global number of posts in different sentiments.

Figure 12. Two sides' stakeholders' number of posts in different sentiments.

6. Discussions, Conclusions, and Limitations

This research has identified various online class topics from Reddit posts. As a daily activity, educators and learners discuss their experiences, including funny and embarrassing incidents, comfortable aspects, school life memories, and creative memes inspired by the class. Therefore, the overall sentiment is neutral. However, sentiments are strong and polarized [33] when diving into posts focused on online classes. Learners' sentiment tends to be negative, expressing a sense of depression and self-doubt, which reflects approval-seeking behaviors. Educators' sentiment is much more rational and positive in contrast. Diverse informal, self-directed learning activities beyond schools' requirements are identified in relevant subreddits, revealing educators' and learners' different motivations for social media usage [8]. This study provides hints to improve online classes to fulfill the needs of educators and learners and how social media can aid online education, as highlighted below.

Results have reflected that cheating in online classes is alarming as a systematic and social issue. Educators should propose solutions on various aspects to lower cheating motivation by cultivating a sense of guilt to combat stimulation from others. Particularly, academic libraries should include related ethics and academic integrity in library instruction [34]. Concerning curriculum design, educators may increase the weight of the grade for assignments in online classes and continuous assessment to decrease the effect of exam cheating on overall course assessment. For example, most subjects of the information management degree (both undergraduate and postgraduate) at The University of Hong Kong have no exams and adopt continuous assessment, group projects, and individual essays [35]. Using similarity-checking tools to detect possible cheating issues through submitted assignments and examination scripts also helps deter cheating through plagiarism [36]. Education and human–computer interaction researchers may work on dealing with the online cheating issue to make the online mode more assessable in formal teaching.

In addition, time series analysis posts doubt the effect of strategies for solving the problems identified in online classes. The decreasing number of posts about online classes is mainly due to the less frequent discussion of this topic among learners because of the fewer online classes. This suggests a longer tracking period is necessary. Moreover, we observed that educators' and learners' post number fluctuations shared a similar pattern, which may indicate that educators' posts are crucial in facilitating discussion. Therefore, we suggest that educators post more posts to encourage the learners to exchange ideas and eventually form a community of practice to engage learners [37].

This research has revealed the concerns behind online classes and verified the different motivations and behavior of social media usage between educators and learners. The

research shows the potential of using Reddit as a data source in education research and reference for follow-up research. Yet, a limitation is that this research could not fully utilize the special structure of Reddit submission and the corresponding meta-data. Future efforts are necessary to consider more attributions of submissions when analyzing Reddit data. Therefore, for future research direction, we plan to use more advanced machine-learning and deep-learning sentiment analysis algorithms to avoid analyzing keywords out of context and further investigate related issues, such as improving the accuracy of the sentiment ratings.

Appendix A

Table A1. Examples of Posts about Cheating.

Example Posts from Educators
It makes me sad because I'm passionate about the subject and love teaching it. Before, people would come to my office, and we would work stuff out. Now, nobody goes on my zoom. Instead, students **cheat** and never get the help they need from me.
Annoyed about people cheating? Develop tests and assessments that you can't **cheat** in or are pointless to **cheat** in. Many of us have done it, including many STEM instructors. Some of you post edgy comments and posts about how excited you get from busting dozens of students using online resources for online tests . . . ? Do you realize that practically the entire class can cheat on straightforward questions? Here's a surprise: everyone is. So change your testing style and stop making the conditions ripe to ruin dozens of lives.
Should I: 1. give all students an opportunity to retake the test, so all have the same chances to cheat? 2. Only let the students who missed the exam take the test by looking at a PDF online and taking it on separate paper with cameras on, so hopefully, I can prevent at least some of the worst teaching. Option 3. Design a new test thru edulastic (an electronic testing program) and have everyone take the new test. Help . . . need advice. BTW . . . yes, **cheating** is that bad at my school. But it's a cultural thing. They see it as working the system and not unethical.

Example Posts from Students
Yep, i **cheat** in the online school. This year has been very stressful with covid and some family situations going on. I feel very difficult to focus in class, and I would rather do something important like practice getting better at speed running or playing the trombone. Every time when I do it, I feel ashamed and guilty for what I do, and it never seems to shake off. I really want to learn and not cheat by missing one section and move on to a new section with different things that are more challenging.
He offered to pay me to take the test for him. I said I would never accept money for that and offered to tutor him for free. There are still 2 weeks before the test, and it's easy math. I have a degree in Math and worked my ass off to graduate in the top 10% of my class. I will not take the test for him, but I also hate the idea of him failing and being set back by a whole year. I legitimately believe that he will be fine without ever learning the material, but I can't bring myself to **cheat** for him.
This class has caused me so much suffering and tears throughout the year, and when I realized the opportunity to cheat because of online learning would be right there, I didn't hesitate. Despite being monitored, I had my notes in front of me every single quiz. I just put them on the wall. The more I **cheated**, the more guilty I felt about this. Cheating goes against my morals. All of my friends have been cheating, and I felt that if I didn't cheat, I would be at a disadvantage.

Appendix B

Table A2. Examples of Posts with Unique Sentimental Words.

Unique Word	Example Posts
Panic (Educators)	Since moving the class online, this student has complained to the dean and my chair that I have increased my demands on the class (not true) and that I am responsible for this student developing **panic** attacks Yesterday, another student asked for the link because she registered for a synchronous class. So I'm like, this is weird, and just respond and confirm again that it's not live. So I'm thinking to myself, I HAD to have missed something—so I go back to my emails, and SURE ENOUGH, it says synchronous. CUE MY **PANIC**. I am at home today for Veteran's day, and I just had a **panic** attack trying to lesson plan. I don't know how much longer I can do this. I am tired of online/in-person teaching at the same time. I am tired of the lack of student participation. I am tired of working 14-h days with little to no payoff. I am tired of nausea I get when I even think about work. Please tell me this gets better.
Successful, Awesome (Educators)	So long story short, I have a double period with students, which is great when in person, but online seems wasted. Small groups do not work online, so successful students are droning away on different online programs, and students who fail or don't participate during class get extra time to do their work and still don't finish it anyways. If you had an extra period, what is something you would do with students for 45 min that isn't a complete waste of time and translates well online? I've never seen these particular students so engaged, curious, and excited to come to my class and learn. Yes, my online students aren't engaged, and some are honestly falling behind due to them not engaging or completing the work, but I can only lead a horse to water. I can't force this age group to do anything. I can only provide them with the resources to make them **successful**. Teachers are like surgeons. Sometimes, we lose them, but we can't save them all. Thankfully I took some notes and planned my lessons accordingly, so this semester has been better overall. Even though I'm still experimenting with stuff (a pallet here, a flip grid there), having my lesson plans laid out instead of starting from scratch lets me preview the class in my head and make a few changes to things I believe weren't as **successful**. For anyone on this same journey, I advise you to hang in there: it gets better.
Depress, Dumb (Learners)	Granted, I didn't do some assignments (**depression** hit me very hard this year, and I was anxious about even attending school), but I am trying to make myself go to class. At first, I was glad I could get some extra sleep and didn't have to worry about putting on an uncomfortable uniform. But as the days went by, I realized that I started getting increasingly tired and couldn't get my work done. I hope I'm not alone when I feel **depressed** and tired and nothing gives me the energy to do work. The weekends don't help me get energy anymore. And my fellow students seem to be ok, and they complete all of their homework on time and perfectly, which doesn't help me. I hope others feel the same way and are not lazy leach or parasite. It makes me feel like a POS because I'm going to tutoring, TAs office hours, going to class, and all of that. I am no longer eligible for the grade replacement opportunity too. I'm so frustrated and feel like a failure and **dumb** compared to my peers.
Trust, Laugh (Learners)	Besides that, I would try to attend as many office hours are you can, **trust** me, they can be your savior. Don't give up as well. Online classes will be hard for those who are not used to them and completely despise them for ruining the semester. But keep social distancing, wear a mask, and maybe by spring or summer, we can all return to our campuses. So, I'm gonna have an online meeting with my class tomorrow, and I want to piss my teacher off and make the others **laugh**. Any idea for what I could say to achieve this? E.g., Excuse me, sir, but may I go to the toilet, please? (Quite simple but something like that) So tomorrow will be my first online class, and no one in it knows how to do background stuff, so please, can someone suggests a good-looking or funny background? It can be a video as well. Let it be funny, so I can do some cool stuff with it or something that would make me look interesting or at least make someone **laugh** (i will make an update if someone suggests it). Whenever I use Reddit during online classes, the memes are super funny, and I cannot control my **laugh**, but other times it is not funny as it when during online class

References

1. Yu, P.Y.; Lam, E.T.H.; Chiu, D.K.W. Operation management of academic libraries in Hong Kong under COVID-19. *Libr. Hi Tech.* **2022**, in press. [CrossRef]
2. Huang, P.S.; Paulino, Y.; So, S.; Chiu, D.K.W.; Ho, K.K.W. Editorial—COVID-19 Pandemic and Health Informatics (Part 1). *Libr. Hi Tech.* **2021**, *39*, 693–695. [CrossRef]
3. Huang, P.-S.; Paulino, Y.C.; So, S.; Chiu, D.K.; Ho, K.K. Guest editorial: COVID-19 Pandemic and Health Informatics Part 2. *Libr. Hi Tech.* **2022**, *40*, 281–285. [CrossRef]
4. Tse, H.L.; Chiu, D.K.W.; Lam, A.H. From reading promotion to digital literacy: An analysis of digitalizing mobile library services with the 5E Instructional Model. In *Modern Reading Practices and Collaboration Between Schools, Family, and Community*; Almeida, A., Esteves, S., Eds.; IGI Global: Hershey, PA, USA, 2022; pp. 239–256.
5. Pokhrel, S.; Chhetri, R. A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning. *High. Educ. Futur.* **2021**, *8*, 133–141. [CrossRef]
6. Yao, L.; Lei, J.; Chiu, D.K.W.; Xie, Z. Adult Learners' Perception of Online Language English Learning Platforms in China. In *New Approaches to the Investigation of Language Teaching and Literature*; Garcés-Manzanera, A., García, M.E.C., Eds.; IGI Global: Hershey, PA, USA, 2023.
7. Petrie, C. Spotlight: Quality Education for All during COVID-19 Crisis | Unesco IIEP Learning Portal. 2020. Available online: https://learningportal.iiep.unesco.org/en/library/spotlight-quality-education-for-all-during-covid-19-crisis (accessed on 10 August 2022).
8. Dong, G.; Chiu, D.K.; Huang, P.-S.; Ho, K.K.; Lung, M.M.-W.; Geng, Y. Relationships between research supervisors and students from coursework-based master's degrees: Information usage under social media. *Inf. Discov. Deliv.* **2021**, *49*, 319–327. [CrossRef]
9. Doucet, A.; Deborah, N.; Koen, T.; Jim Francis, T. Thinking about Pedagogy in an Unfolding Pandemic by Education International—Issuu. 2020. Available online: https://issuu.com/educationinternational/docs/2020_research_covid-19_eng (accessed on 10 August 2022).
10. Cheng, J.; Yuen, A.H.; Chiu, D.K. Systematic review of MOOC research in mainland China. *Libr. Hi Tech.* **2022**. ahead-of-print. [CrossRef]
11. Kulal, A.; Nayak, A. A study on perception of teachers and students toward online classes in Dakshina Kannada and Udupi District. *Asian Assoc. Open Univ. J.* **2020**, *15*, 285–296. [CrossRef]
12. Zia, A. Exploring factors influencing online classes due to social distancing in COVID-19 pandemic: A business students perspective. *Int. J. Inf. Learn. Technol.* **2020**, *37*, 197–211. [CrossRef]
13. Sarkar, S.S.; Das, P.; Rahman, M.M.; Zobaer, M.S. Perceptions of Public University Students Towards Online Classes During COVID-19 Pandemic in Bangladesh. *Front. Educ.* **2021**, *6*, 703723. [CrossRef]
14. Ye, S.; Ho, K.K.W. College students' Twitter usage and psychological well-being from the perspective of generalised trust: Comparing changes before and during the COVID-19 pandemic. *Libr. Hi Tech.* **2022**, in press. [CrossRef]
15. Leung, T.N.; Hui, Y.M.; Luk, C.K.; Chiu, D.K.; Ho, K.K. Evaluating Facebook as aids for learning Japanese: Learners' perspectives. *Libr. Hi Tech.* **2022**, in press. [CrossRef]
16. Xie, R.; Chu, S.K.W.; Chiu, D.K.W.; Wang, Y. Exploring Public Response to COVID-19 on Weibo with LDA Topic Modeling and Sentiment Analysis. *Data Inf. Manag.* **2021**, *5*, 86–99. [CrossRef] [PubMed]
17. Fang, J.; Chiu, D.K.W.; Ho, K.K.W. *Exploring Cryptocurrency Sentimental with Clustering Text Mining on Social Media, In Handbook of Research on Intelligent Analytics with Multi-Industry Applications*; Sun, Z., Ed.; IGI Global: Hershey, PA, USA, 2021; pp. 157–171.
18. He, Z.; Chiu, D.K.W.; Ho, K.K.W. Weibo Analysis on Chinese Cultural Knowledge for Gaming. In *Handbook of Research on Foundations and Applications of Intelligent Business Analytics*; Sun, Z., Ed.; IGI Global: Hershey, PA, USA, 2022; pp. 320–349.
19. Klebanov, S. Reddit's Journey to a Potential $15 Billion IPO. 2022. Available online: https://www.businessofbusiness.com/articles/reddit-15-billion-ipo-cultural-force/ (accessed on 12 August 2022).
20. Todorov, G. 70+ Important Reddit Statistics 2022. 2021. Available online: https://thrivemyway.com/reddit-statistics/ (accessed on 10 August 2022).
21. Proferes, N.; Jones, N.; Gilbert, S.; Fiesler, C.; Zimmer, M. Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Soc. Media + Soc.* **2021**, *7*, 20563051211019004. [CrossRef]
22. Willet, K.B.S.; Carpenter, J.P. Teachers on Reddit? Exploring contributions and interactions in four teaching-related subreddits. *J. Res. Technol. Educ.* **2020**, *52*, 216–233. [CrossRef]
23. Jamnik, M.; Lane, D. The use of Reddit as Gifted education for high-quality data. *Pract. Assess. Res. Eval.* **2019**, *22*, 5. [CrossRef]
24. Hodges, J.; Simonsen, M.; Ottwein, J.K. Gifted Education on Reddit: A Social Media Sentiment Analysis. *Gift. Child Q.* **2022**, *66*, 296–315. [CrossRef]
25. Zirikly, A.; Resnik, P.; Uzuner, Ö.; Hollingshead, K. CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, Minneapolis, MN, USA, 6 June 2019; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 24–33.
26. Ammari, T.; Schoenebeck, S.; Romero, D. Self-declared throwaway accounts on Reddit: How platform affordances and shared norms enable parenting disclosure and support. *Proc. ACM Hum.-Comput. Interact.* **2019**, *3*, 135:1–135:30. [CrossRef]

27. Kovacs, E.-R.; Cotfas, L.-A.; Delcea, C. COVID-19 Vaccination Opinions in Education-Related Tweets. Paper presented at the Eurasian Business and Economics Perspectives. In Proceedings of the 37th Eurasia Business and Economics Society Conference, Berlin, Germany, 6–8 October 2021; Springer International Publishing: Cham, Switzerland, 2021; pp. 21–41.

28. Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; Blackburn, J. The pushshift Reddit dataset. *Proc. Int. AAAI Conf. Web Soc. Media* **2020**, *14*, 830–839. [CrossRef]

29. Gaffney, D.; Matias, J.N. Caveat emptor, computational social science: Large-scale missing data in a widely-published Reddit corpus. *PLoS ONE* **2018**, *13*, e0200162. [CrossRef]

30. Hussein, D.M.E.-D.M. A survey on sentiment analysis challenges. *J. King Saud Univ. Eng. Sci.* **2018**, *30*, 330–338. [CrossRef]

31. Hutto, C.; Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proc. Int. AAAI Conf. Web Soc. Media* **2014**, *8*, 216–225. [CrossRef]

32. Moore, C.; Chuang, L. Redditors revealed: Motivational factors of the Reddit community. In Proceedings of the Hawaii International Conference on System Sciences 2017 (HICSS-50), Hilton Waikoloa Village, HI, USA, 4–7 January 2017; Available online: https://aisel.aisnet.org/hicss-50/dsm/dsm_and_communities/7 (accessed on 12 August 2022).

33. Au, C.H.; Ho, K.K.W.; Chiu, D.K.W. Does political extremity harm the ability to identify online information validity? Testing the impact of polarisation. *Gov. Inf. Q.* **2021**, *38*, 101602. [CrossRef]

34. Lomness, A.; Lacey, S.; Brobbel, A.; Freeman, T. Seizing the opportunity: Collaborative creation of academic integrity and information literacy LMS modules for undergraduate Chemistry. *J. Acad. Libr.* **2021**, *47*, 102328. [CrossRef]

35. Yew, A.; Chiu, D.K.W.; Nakamura, Y.; Li, K.K. Quantitative Comparison of LIS Programs Accredited by ALA and CILIP. *Libr. Hi Tech.* **2022**, *40*, 1721–1745. [CrossRef]

36. Cleophas, C.; Hönnige, C.; Meisel, F.; Meyer, P. Who's Cheating? Mining Patterns of Collusion from Text and Events in Online Exams. *INFORMS Trans. Educ.* **2023**, *23*, 84–94. [CrossRef]

37. Lei, S.Y.; Chiu, D.K.; Lung, M.M.-W.; Chan, C.T. Exploring the aids of social media for musical instrument education. *Int. J. Music. Educ.* **2021**, *39*, 187–201. [CrossRef]